

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Requestor's
Name: _____

Serial
Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 08-23-02

Searcher: Beverly C 4954

Terminal time: 20

Elapsed time: _____

CPU time: _____

Total time: 40

Number of Searches: _____

Number of Databases: 1

Search Site

_____ STIC

_____ CM-1

_____ Pre-S

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

☒ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:01:00 ; Search time 179.72 Seconds
(without alignments)
5.562 Million cell updates/sec

Title: US-08-854-825-34
Perfect score: 40
Sequence: 1 SLMAFTAAV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```

Database :
A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Description		
	Score	Match	Length DB	ID		
1	40	100.0	9	16	AAR84575	Cytotoxic T-cell e
2	40	100.0	9	18	AAW39561	HCV1 NS4 peptide (
3	40	100.0	9	20	AAAY10234	T cell epitope/MHC
4	40	100.0	9	20	AAAY10512	HLA Class I motif
5	40	100.0	13	22	AAB82061	HLA Class I motif
6	40	100.0	20	22	AAB82060	HCV antigen, NS4 1
7	40	100.0	117	10	AAAP90136	HCV antigen, NS4 1
8	40	100.0	117	10	AAAP92019	Sequence of hepati
9	40	100.0	117	14	AAR33872	Sequence of hepati
10	40	100.0	118	12	AAAR13354	Polypeptide p1689
11	40	100.0	128	10	AAAP90135	p1689 HCV antigen
						Sequence of hepati

12	40	100.0	128	10	AAP92018	Sequence of the po
13	40	100.0	237	10	AAP90138	Peptide encoded by
14	40	100.0	237	10	AAP92021	Polypeptide encode
15	40	100.0	269	20	AAR92815	HCV NS4B protein.
16	40	100.0	363	13	AAR23999	Open reading frame
17	40	100.0	363	17	AAR90933	HCV antigen C100.
18	40	100.0	382	10	AAP90182	C terminus of supe
19	40	100.0	382	10	AAP92048	Sequence of the ca
20	40	100.0	460	10	AAP90141	Protein sequence o
21	40	100.0	460	10	AAP92024	Polypeptide encode
22	40	100.0	512	22	AAB69031	HCV recombinant an
23	40	100.0	592	14	AAR33565	CXS-HCV antigen fu
24	40	100.0	592	22	AAB89023	HCV recombinant an
25	40	100.0	594	14	AAR33566	CXS-HCV antigen fu
26	40	100.0	594	22	AAB69024	HCV recombinant an
27	40	100.0	597	13	AAR21571	HCV CXS-C100D1 - p
28	40	100.0	597	14	AAR33638	HCV C100D1 recombi
29	40	100.0	597	14	AAR33580	HCV-C100D1 recombi
30	40	100.0	597	14	AAR33600	HCV C100D1 recombi
31	40	100.0	597	22	AAB51378	HCV recombinant an
32	40	100.0	599	13	AAR21572	HCV CXS-C100D2 - p
33	40	100.0	599	14	AAR33639	HCV C100D2 recombi
34	40	100.0	599	14	AAR33581	HCV-C100D2 recombi
35	40	100.0	599	14	AAR33601	HCV C100D2 recombi
36	40	100.0	599	22	AAB51379	HCV recombinant an
37	40	100.0	613	14	AAR33567	CXS-HCV antigen fu
38	40	100.0	613	22	AAB69025	HCV recombinant an
39	40	100.0	781	13	AAR21565	HCV CXS-33-BCD - p
40	40	100.0	781	13	AAR22208	Sequence of fusion
41	40	100.0	781	14	AAR33632	HCV CXS-33-BCD fus
42	40	100.0	781	14	AAR33574	HCV CXS-33-BCD pro
43	40	100.0	781	14	AAR33594	HCV CXS-33-BCD pro
44	40	100.0	781	15	AAR52690	HCV CXS-33-BCD ant
45	40	100.0	781	22	AAB51372	HCV recombinant an

ALIGNMENTS

RESULT	1
AAR84575	
ID	AAR84575 standard; peptide: 9 AA.
XX	AC
XX	AAR84575;
XX	
XX	25-APR-1996 (first entry)
XX	
XX	
DE	Cytotoxic T-cell epitope, aa 1789-1797 of HCV-1 NS4 region.
XX	
KW	Hepatitis C virus; HCV; epitope; vaccine; immunogen.

CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
 CC were tested for peptide specific cytotoxic T-cell activity. The
 CC peptides AAR84570-77 were found to have substantial homology with a T-
 CC cell epitope and are useful in vaccines against HCV infection.

SQ Sequence 9 AA;

Query Match 100.0%; Score 40; DB 16; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
 |||||
 DB 1 simaftaav 9

RESULT 2

AAW39561
 ID AAW39561 standard; peptide; 9 AA.

AC AAW39561;

DT 11-JUN-1998 (first entry)

DE HCV1 NS4 peptide (pos.1789-1797) capable of binding to HLA-A*0201.

XX T cell epitope; immune response; human leukocyte antigen; HLA Class I;
 KW vaccine; immunogenic; major histocompatibility complex; MHC; B cell;
 KW disease; anti-tumour; anti-viral.

OS Hepatitis C virus.

XX WO9741440-A1.

PN 06-NOV-1997.

PD 28-APR-1997; 97WO-NL00229.

XX 23-DEC-1996; 96EP-0203670.

PR 26-APR-1996; 96EP-0201145.

XX (UYLE-) RIJKSUNIV LEIDEN.

PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.

XX Kast WM, Mellief CJM, Offringa R, Toes REM, Van Der Burg SH;

PI WPI; 1997-549891/50.

XX Method of selecting T cell peptide epitope(s) - by measuring the
 PT stability of HLA class I-peptide complexes on intact B cells

PS Example 2; Page 66; 109pp; English.

XX Peptides AAW39430-W39734 are used in a novel method for the selection of
 CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
 CC method involves the identification of peptide sequences capable of
 CC binding to an HLA (human leukocyte antigen) class I molecule and
 CC measuring the binding of this epitope peptide to the HLA class I
 CC peptide. The stability of binding of the peptide and MHC (major
 CC histocompatibility complex) class I molecule is measured on intact human
 CC B cells carrying the MHC molecule at their cell surfaces. The method can
 CC be used to select peptide epitopes for generating vaccines against a
 CC disease associated with the polypeptide, e.g. cancers or AIDS. The
 CC peptide epitopes are especially T-cell peptide epitopes with strong
 CC anti-tumour and anti-viral immune responses. Peptide AAW39561 is a
 CC conserved hepatitis C virus type 1 NS4 derived peptide used to determine
 CC the stability of HLA-A*0201 complexed with known CTL lymphocytes.

SQ Sequence 9 AA;

Query Match 100.0%; Score 40; DB 18; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
 |||||
 DB 1 simaftaav 9

RESULT 3

AAV10234
 ID AAV10234 standard; Peptide; 9 AA.

XX AAV10234;

DT 12-MAY-1999 (first entry)

DE T cell epitope/MHC ligand SEQ ID NO:164.

XX Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX Synthetic.

OS Hepatitis C virus.

XX WO9902183-A2.

PN 21-JAN-1999.

XX 10-JUL-1998; 98WO-US14289.

PR 10-DEC-1997; 97US-0988320.

PR 10-JUL-1997; 97CA-2209815.

XX (CTLI-) CTL IMMUNOTHERAPIES CORP.

PA Kuendig TM, Simard JLL;

PI WPI; 1999-120514/10.

DR Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX Disclosure: Page 30; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAV10071 to AAV10639 represent examples of peptide
 CC antigens given in the present invention.

SQ Sequence 9 AA;

Query Match 100.0%; Score 40; DB 20; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
 |||||
 DB 1 simaftaav 9

DT	22-JUN-2001	(first entry)	
XX			
DE	HCV antigen, NS4 1787-1799.		
XX			
KW	Antigen; immunostimulant; vaccine; pharmaceutical composition; antiviral;		
KW	viral infection.		
XX			
OS	Hepatitis c virus.		
XX			
PN	WO200124822-A2.		
XX			
PD	12-APR-2001.		
XX			
XX	02-OCT-2000; 2000WO-EP09657.		
PF			
XX			
PR	01-OCT-1999; 99AT-0001680.		
XX			
PA	(CIST-) CISTEM BIOTECHNOLOGIES GMBH.		
XX			
PI	Fleitmann J, Mattner F, Buschle M, Melling J;		
XX			
DR	WPI; 2001-290577/30.		
XX			
PT	New pharmaceutical composition comprising an antigen, an		
PT	immunostimulating substance and a polycationic polymer, useful in		
PT	manufacturing vaccines -		
XX			
PS	Claim 12; Page 16; 20pp; English.		
XX			
CC	The present invention relates to a pharmaceutical composition comprising		
CC	(a) an antigen; (b) an immunostimulating substance consisting of		
CC	neuroactive compounds, hormones, compounds having growth hormone activity		
CC	or their mixtures; and (c) a polycationic polymer. The present sequence		
CC	is an antigenic peptide derived from Hepatitis c virus, which was used in		
CC	the present invention. The composition is useful in manufacturing		
CC	vaccines.		
XX			
SQ	Sequence 13 AA;		
Query Match 100.0%; Score 40; DB 22; Length 13;			
Best Local Similarity 100.0%, Pred. No. 0.027;			
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0			
Qy	1 SLMAFTAAV 9		
Db	3 slmaftaav 11		
RESULT 6			
AAB82060			
ID	AAB82060 standard; peptide; 20 AA.		
XX			
AC	AAB82060;		
XX			
DT	22-JUN-2001 (first entry)		
XX			
DE	HCV antigen, NS4 1785-1804.		
XX			
KW	Antigen; immunostimulant; vaccine; pharmaceutical composition; antiviral;		
KW	viral infection.		
XX			
OS	Hepatitis c virus.		
XX			
PN	WO200124822-A2.		
XX			
PD	12-APR-2001.		
XX			
PF	02-OCT-2000; 2000WO-EP09657.		
XX			
PR	01-OCT-1999; 99AT-0001680.		

XX
PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI Fleitmann J, Mattner F, Buschle M, Melling J;
XX
XX WPI; 2001-290577/30.
DR
XX
XX New pharmaceutical composition comprising an antigen, an
PT immunostimulating substance and a polycationic polymer, useful in
PT manufacturing vaccines -
XX
XX
PS Claim 12; Page 16; 20pp; English.
XX
XX The present invention relates to a pharmaceutical composition comprising
CC (a) an antigen; (b) an immunostimulating substance consisting of
CC neuroactive compounds, hormones, compounds having growth hormone activity
CC or their mixtures; and (c) a polycationic polymer. The present sequence
CC is an antigenic peptide derived from Hepatitis C virus, which was used in
CC the present invention. The composition is useful in manufacturing
CC vaccines.
XX
XX Sequence 20 AA;
SQ

Query Match 100.0%; Score 40; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 5 slmaftaav 13
|||||

RESULT 7
AAP90136
ID AAP90136 standard; protein; 117 AA.
XX
AC AAP90136;
XX
DT 01-NOV-1989 (first entry)
XX
DE Sequence of hepatitis C virus cDNA insert in DNA 81.
XX
XX Hepatitis C virus; DNA 81; probe; vaccine.
XX
XX Pan troglodytes.
XX
XX GB2212511-A.
XX
PD 26-JUL-1989.
XX
XX 18-NOV-1988; 88GB-0027024.
XX
XX 18-NOV-1987; 87US-0122714.
XX
PA (CHIR) CHIRON CORPORATION.
XX
PI Houghton M, Choo QL, Kuo G;
XX
XX WPI; 1989-215054/30.
DR N-PSDB; AAN90305.
XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.
XX
XX Disclosure; fig 4; 235pp; English.
XX
XX The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in DNA 81 (see AAN90305). The polypeptides are
CC used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.
XX

SQ Sequence 117 AA;

Query Match 100.0%; Score 40; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 101 slmaftaav 109
|||||

RESULT 8
AAP92019
ID AAP92019 standard; protein; 117 AA.
XX
AC AAP92019;
XX
DT 02-MAR-1990 (first entry)
XX
DE Sequence of the polypeptide encoded in the hepatitis C virus (HCV) cDNA
DE insert in clone 81.
XX
XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
KW
XX
OS Hepatitis C virus.
XX
PN EP318216-A.
XX
PD 31-MAY-1989.
XX
XX 18-NOV-1988; 88EP-0310922.
XX
XX 14-NOV-1988; 88US-0271450; US-122714.
XX
PA (CHIR) CHIRON CORP.
XX
PI Houghton M, Choo q-L, Kuo G;
XX
XX WPI; 1989-159274/22.
DR N-PSDB; AAN92075.
XX
XX Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX
XX Claim 13; Figure 4; 139pp; English.
XX
XX It is the putative sequence encoded in the open reading frame of
CC hepatitis C virus (HCV) cDNA insert in clone 81. It is an epitope,
CC portions of which could be used as immunoassay reagents and vaccines and
CC to generate antibodies useful in diagnosis and passive immunotherapy for
CC HCV infection/non-A, non-B hepatitis.
XX
XX
SQ Sequence 117 AA;

Query Match 100.0%; Score 40; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 101 slmaftaav 109
|||||

RESULT 9
AAR33872
ID AAR33872 standard; peptide; 117 AA.
XX
AC AAR33872;
XX
XX 19-JUL-1993 (first entry)
DT
XX

DE Polypeptide p1689 comprising HCV viral antigen.

XX Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;
KW p408.

XX Synthetic.

XX WO9306247-A.

XX 01-APR-1993.

XX 16-SEP-1992; 92WO-US07813.

XX 16-SEP-1991; 91US-0760292.

XX (ABBO) ABBOTT LAB.

XX Lesniewski RR, Leung TK;

XX WPI; 1993-117563/14.

XX Assay for detecting presence of antibody to hepatitis C viral
PT antigen - by contacting sample with polypeptide contg. at least
PT one epitope of virus antigen

XX Disclosure; Page 13; 63pp; English.

XX The synthetic peptide p1689 represents amino acid residues 1689-1805 of
CC the hepatitis C viral antigen. The peptide may be used in an assay to
CC detect antibodies to HCV and thus to diagnose chronic HCV infection.
CC See also AAR33861-87.

XX Sequence 117 AA;

Query Match 100.0%; Score 40; DB 14; Length 117;

Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

Db 101 slmaftaav 109

RESULT 10

AAR13354

ID AAR13354 standard; Protein; 118 AA.

XX AAR13354;

XX 23-OCT-1991 (first entry)

XX p1689 HCV antigen (1689-1805).

XX C100-3; hepatitis C virus; immunoassay; epitope.

XX Synthetic.

XX AU068390-A.

XX 27-JUN-1991.

XX 21-DEC-1990; 90AU-0068390.

XX 07-NOV-1990; 90US-0610180.

XX 22-DEC-1989; 89US-0456162.

XX (ABBO) ABBOTT LABORATORIES.

XX WPI; 1991-238393/33.

XX Immunological assays for hepatitis C virus antibody - by using
PT polypeptide(s) contg. epitope(s) of hepatitis C virus antigens

XX Claim 1; Page 45; 62pp; English.

XX The polypeptide may be prepared by solid phase synthesis fragment
CC coupling (pref.) or using recombinant technology.

XX The assay has increased sensitivity and is more specific than
CC assays using the polypeptide C100-3 (EP-318216).

XX See also AAQ13146-48 and AAR13343-65.

XX Sequence 118 AA;

Query Match 100.0%; Score 40; DB 12; Length 118;

Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

Db 102 slmaftaav 110

RESULT 11

AAP90135

ID AAP90135 standard; protein; 128 AA.

XX AAP90135;

XX 01-NOV-1989 (first entry)

XX Sequence of hepatitis C virus cDNA insert in clone 5-1-1, 81
DE 91 and 1-2.

XX Hepatitis C virus; clone 5-1-1, 81, 91, 1-2; probe; vaccine.

XX Pan troglodytes.

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-215054/30.

XX N-PSDB; AAN90304.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.

XX Disclosure; fig 3; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in clone 5-1-1, 81, 91 and 1-2 (see AAN90304).

XX The polypeptides are used to diagnose HCV-induced NANBH,

XX to raise antibodies for immunoassay or treatment, or to
CC produce vaccines.

XX Sequence 128 AA;

Query Match 100.0%; Score 40; DB 10; Length 128;

Best Local Similarity 100.0%; Pred. No. 0.29;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

Db 112 slmaftaav 120

```
RESULT 12
AAP92018
ID AAP92018 standard; protein; 128 AA.
XX
AC AAP92018;
XX
DT 02-MAR-1990 (first entry)
XX
DE Sequence of the polypeptide encoded in a composite sequence of the
CC hepatitis C virus (HCV) cDNA.
XX
KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HAMBH)
XX
OS Hepatitis C virus.
XX
PN EP318216-A.
XX
PD 31-MAY-1989.
XX
PF 18-NOV-1988; 88EP-0310922.
XX
PR 14-NOV-1988; 88US-0271450; US-122714.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Houghton M, Choo q-L, Kuo G;
XX
DR WPI; 1989-159274/22.
DR N-PSDB; AAN92074.
XX
PT Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX
PS Claim 13; Figure 3; 139pp; English.
XX
CC It is encoded in the open reading frame of a composite nucleotide
CC sequence derived from overlapping hepatitis C virus (HCV) cDNA in clones
CC 81, 1-2, and 91, isolated using a synthetic sequence equivalent to a
CC fragment of HCV cDNA in clone 5-1-1. It is an epitope, portions of which
CC could be used as immunoassay reagents and vaccines and to generate
CC antibodies useful in diagnosis and passive immunotherapy for HCV
CC infection/non-A, non-B hepatitis.
XX
SQ Sequence 128 AA;

Query Match 100.0%; Score 40; DB 10; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 112 slmaftaav 120

RESULT 13
AAP90138
ID AAP90138 standard; protein; 237 AA.
XX
AC AAP90138;
XX
DT 01-NOV-1989 (first entry)
XX
DE Peptide encoded by combined ORF of hepatitis C virus cDNAs in
DE clones 36 and 81.
XX
KW Hepatitis C virus; open reading frame; clone 36; clone 81; probe;
KW vaccine.
XX
OS Pan troglodytes.
XX
```

```
PN GB2212511-A.
XX
PD 26-JUL-1989.
XX
PF 18-NOV-1988; 88GB-0027024.
XX
PR 18-NOV-1987; 87US-0122714.
XX
PA (CHIR ) CHIRON CORPORATION.
XX
PI Houghton M, Choo QL, Kuo G;
XX
DR WPI; 1989-215054/30.
DR N-PSDB; AAN90307.
XX
PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX
PS Disclosure; fig 6; 235pp; English.
XX
CC The sequence is the peptide encoded by the combined open reading frame
CC of hepatitis C virus (HCV) cDNAs in clones 36 and 81 (see AAN90307).
CC The polypeptides are used to diagnose HCV-induced NANBH, to raise
CC antibodies for immunoassay or treatment, or to produce vaccines.
XX
SQ Sequence 237 AA;

Query Match 100.0%; Score 40; DB 10; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 221 slmaftaav 229

RESULT 14
AAP92021
ID AAP92021 standard; protein; 237 AA.
XX
AC AAP92021;
XX
DT 02-MAR-1990 (first entry)
XX
DE Polypeptide encoded in the combined open reading frames of clones 36 and
DE 81 of hepatitis C virus (HCV) cDNA.
XX
KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HAMBH)
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT misc_feature 1..135 /*tag= a
FT misc_feature 121..237 /*tag= b
FT misc_feature 121..237 /*tag= c
XX
PN EP318216-A.
XX
PD 31-MAY-1989.
XX
PF 18-NOV-1988; 88EP-0310922.
XX
PR 14-NOV-1988; 88US-0271450; US-122714.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Houghton M, Choo q-L, Kuo G;
XX
```


DR WPI; 1989-159274/22.
XX N-PSDB; AAN92077.
PT Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX Claim 13; Figure 6; 139pp; English.
PS
XX
CC It is the polypeptide sequence encoded in the combined open reading
CC frames of the hepatitis C virus (HCV) cDNA of clones 36 and 81. It is not
CC necessarily physically derived from HCV cDNA but may be generated in any
CC manner. Tag a = clone 36. Tag b = clone 81. Tag c = the region of overlap
CC between clones 36 and 81. It is an epitope, portions of which could be
CC used as immunoassay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
XX
SQ Sequence 237 AA;

Query Match 100.0%; Score 40; DB 10; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 221 slmaftaav 229
|||||

RESULT 15
AAW92815
ID AAW92815 standard; protein; 269 AA.
XX
XX AC AAW92815;
XX
DT 10-MAY-1999 (first entry)
XX
DE HCV NS4B protein.
XX
KW NS4B protein; ATPase; Flaviviridae; antiviral compound; Dengue virus;
KW therapeutic agent; yellow fever virus; pestiviruses; swine fever;
KW bovine viral diarrhoea virus.
XX
OS Hepatitis C virus.
XX
PN WO9901582-A1.
XX
PD 14-JAN-1999.
XX
PF 01-JUL-1998; 98WO-US13790.
XX
PR 02-JUL-1997; 97US-0051582.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Delvecchio A, Zhong W;
XX
DR WPI; 1999-106080/09.
XX
XX
PT Detecting Hepatitis C Virus NS4B protein modulators - useful to
PT treat infection with viruses of the Flaviviridae family
XX
PS Claim 6; Page 21-22; 27pp; English.
XX
XX This sequence is used in a novel method for identifying a compound that
CC alters activity of Hepatitis C Virus (HCV) NS4B protein. HCV NS4B
CC modulators can be used as antiviral compounds and as therapeutic agents
CC to treat viruses of the Flaviviridae family, including HCV, yellow fever
CC virus, Dengue viruses types 1-4, and pestiviruses such as bovine viral
CC diarrhoea virus and classic swine fever. Treatment of an HCV-infected
CC human with a HCV1 NS4B antagonist, and treatment of an HCV-infected
CC mammal with a HCV NS4B agonist is claimed.
XX

SQ Sequence 269 AA;
Query Match 100.0%; Score 40; DB 20; Length 269;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 78 slmaftaav 86
|||||

Search completed: August 23, 2002, 10:01:00
Job time: 390 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:52 ; Search time 61.77 Seconds
(without alignments)
3.559 Million cell updates/sec

Title: US-08-854-825-34
Perfect score: 40
Sequence: 1 SLMAFTAAV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	9	1	US-08-214-650-34
2	40	100.0	117	4	US-08-444-818-10
3	40	100.0	117	5	PCT-US92-07813-12
4	40	100.0	128	4	US-08-444-818-8
5	40	100.0	237	4	US-08-444-818-14
6	40	100.0	269	3	US-09-100-557-1
7	40	100.0	382	4	US-08-444-818-68
8	40	100.0	460	4	US-08-444-818-20
9	40	100.0	512	4	US-08-867-611-58
10	40	100.0	592	4	US-08-867-611-47
11	40	100.0	594	4	US-08-867-611-47
12	40	100.0	597	4	US-08-867-611-16
13	40	100.0	597	5	PCT-US92-06965A-21
14	40	100.0	599	4	US-08-867-611-18
15	40	100.0	599	5	PCT-US92-06965A-23
16	40	100.0	613	4	US-08-867-611-49
17	40	100.0	739	4	US-08-444-818-148
18	40	100.0	781	4	US-08-867-611-4
19	40	100.0	781	5	PCT-US92-06965A-9
20	40	100.0	859	4	US-08-444-818-30
21	40	100.0	971	4	US-08-867-611-52
22	40	100.0	973	4	US-08-867-611-53
23	40	100.0	992	4	US-08-867-611-54
24	40	100.0	1021	1	US-07-910-760-12
25	40	100.0	1021	1	US-08-440-519-12
26	40	100.0	1021	4	US-08-440-549-12
27	40	100.0	1786	4	US-08-444-818-54

28	40	100.0	2261	4	US-08-444-818-66	Sequence 66, Appl
29	40	100.0	2436	4	US-08-444-818-75	Sequence 75, Appl
30	40	100.0	2772	4	US-08-444-818-89	Sequence 89, Appl
31	40	100.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
32	40	100.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
33	40	100.0	2894	3	US-08-467-902A-23	Sequence 23, Appl
34	40	100.0	2894	4	US-09-275-265-23	Sequence 23, Appl
35	40	100.0	2955	2	US-08-443-260-3	Sequence 3, Appl
36	40	100.0	2955	3	US-08-442-805A-3	Sequence 3, Appl
37	40	100.0	2955	3	US-08-443-900A-3	Sequence 3, Appl
38	40	100.0	2955	4	US-08-444-818-124	Sequence 124, App
39	40	100.0	2955	4	US-08-249-843-3	Sequence 3, Appl
40	40	100.0	2995	4	US-08-444-818-138	Sequence 138, App
41	40	100.0	3011	1	US-08-188-281B-1	Sequence 1, Appl
42	40	100.0	3011	1	US-08-453-552-1	Sequence 1, Appl
43	40	100.0	3011	1	US-08-453-552-2	Sequence 2, Appl
44	40	100.0	3011	1	US-08-440-103-36	Sequence 36, Appl
45	40	100.0	3011	1	US-08-440-542-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-34
; Sequence 34, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Ceiny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-214-650-34

Query Match 100.0%; Score 40; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
| | | | | | | | | |

Db 1 SLMAFTAAV 9

RESULT 2

US-08-444-818-10

Sequence 10, Application US/08444818

Patent No. 6150087

GENERAL INFORMATION:

APPLICANT: Chien, David Y.

APPLICANT: Rutter, William J.

TITLE OF INVENTION: NANBV Diagnostics and Vaccines

NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,818

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/403,590

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0110.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876

TELEFAX: (508)359-3885

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-444-818-10

Query Match 100.0%; Score 40; DB 4; Length 117;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9

Db 101 SLMAFTAAV 109

RESULT 3

PCT-US92-07813-12

Sequence 12, Application PC/TUS9207813

GENERAL INFORMATION:

APPLICANT: LESNIEWSKI, RICHARD R.

APPLICANT: LEUNG, TAT K.

TITLE OF INVENTION: HEPATITIS C ASSAY

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES

STREET: ONE ABBOTT PARK ROAD

CITY: ABBOTT PARK

STATE: ILLINOIS

COUNTRY: U.S.A.

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/07813

FILING DATE: 19920916

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: FOREMSKIP, PRISCILLA E.

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 4767.P3.03

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365

TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US92-07813-12

Query Match 100.0%; Score 40; DB 5; Length 117;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9

Db 101 SLMAFTAAV 109

RESULT 4

US-08-444-818-8

Sequence 8, Application US/08444818

Patent No. 6150087

GENERAL INFORMATION:

APPLICANT: Chien, David Y.

APPLICANT: Rutter, William J.

TITLE OF INVENTION: NANBV Diagnostics and Vaccines

NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,818

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/403,590

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0110.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876

TELEFAX: (508)359-3885

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 128 amino acids

TYPE: amino acid

TOPOLOGY: linear

```
; MOLECULE TYPE: protein
; US-08-444-818-8

Query Match      100.0%; Score 40; DB 4; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTRAV 9
Db 112 SLMAFTRAV 120

RESULT 5
US-08-444-818-14
; Sequence 14, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-14

Query Match      100.0%; Score 40; DB 4; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTRAV 9
Db 221 SLMAFTRAV 229

RESULT 6
US-09-100-557-1
; Sequence 1, Application US/09100557
; Patent No. 6010848
; GENERAL INFORMATION:
; APPLICANT: DelVecchio, Alfred
; APPLICANT: Zhong, Weidong

; TITLE OF INVENTION: SCREENING METHODS USING AN
; TITLE OF INVENTION: ATPASE PROTEIN FROM A VIRUS OF THE FLAVIVIRIDAE FAMILY
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/100,557
; APPLICATION NUMBER:
; FILING DATE: 19-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051,582
; FILING DATE: 02-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dinner, Dara L.
; REGISTRATION NUMBER: 33,680
; REFERENCE/DOCKET NUMBER: P50675
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5017
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-100-557-1

Query Match      100.0%; Score 40; DB 3; Length 269;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTRAV 9
Db 78 SLMAFTRAV 86

RESULT 7
US-08-444-818-68
; Sequence 68, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-14
```

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-68

Query Match 100.0%; Score 40; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 235 SLMAFTAAV 243

RESULT 8
US-08-444-818-20
Sequence 20, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-20

Query Match 100.0%; Score 40; DB 4; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 319 SLMAFTAAV 327

RESULT 9
US-08-867-611-58
Sequence 58, Application US/08867611
Patent No. 6172189
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DALEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
TITLE OF INVENTION: ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,611
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,757
FILING DATE:
APPLICATION NUMBER: US/08/179,896
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,822
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,565
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,566
FILING DATE: 21-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKI, PRISCILLA E
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-611-58

Query Match 100.0%; Score 40; DB 4; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

|||||

Db 360 SLMAFTAAV 368

RESULT 10

US-08-867-611-47
Sequence 47, Application US/08867611
Patent No. 6172189

GENERAL INFORMATION:

APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
TITLE OF INVENTION: ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/867,611

FILING DATE: 02-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/646,757

FILING DATE:

FILING DATE:

APPLICATION NUMBER: US/08/179,896

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/572,822

FILING DATE: 24-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/614,069

FILING DATE: 07-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,561

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,565

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,566

FILING DATE: 21-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: FOREMSKI, PRISCILLA E

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 4834.US.P6

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 592 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-611-47

Query Match 100.0%; Score 40; DB 4; Length 592;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

|||||

Db 445 SLMAFTAAV 453

RESULT 11

US-08-867-611-48

Sequence 48, Application US/08867611

Patent No. 6172189

GENERAL INFORMATION:

APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
TITLE OF INVENTION: ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/867,611

FILING DATE: 02-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/646,757

FILING DATE:

FILING DATE:

APPLICATION NUMBER: US/08/179,896

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/572,822

FILING DATE: 24-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/614,069

FILING DATE: 07-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,561

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,565

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,566

;
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 594 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-867-611-48

Query Match 100.0%; Score 40; DB 4; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAABV 9
Db 447 SLMAFTAABV 455

RESULT 12
US-08-867-611-16
; Sequence 16, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561

;
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 597 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-867-611-16

Query Match 100.0%; Score 40; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAABV 9
Db 445 SLMAFTAABV 453

RESULT 13
PCT-US92-06965A-21
; Sequence 21, Application PC/TUS9206965A
; GENERAL INFORMATION:
; APPLICANT: DEVARE, S.
; APPLICANT: DESAI, S.
; APPLICANT: DAILEY, S.
; TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.
; ZIP: 60065-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06965A
; FILING DATE: 19920821
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834PC.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 597 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-06965A-21

Query Match 100.0%; Score 40; DB 5; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 445 SLMAFTAAV 453

RESULT 14
US-08-867-611-18
; Sequence 18, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-867-611-18

Query Match 100.0%; Score 40; DB 4; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 447 SLMAFTAAV 455

RESULT 15
PCT-US92-06965A-23
; Sequence 23, Application PC/TUS9206965A
; GENERAL INFORMATION:
; APPLICANT: DEVARE, S.
; APPLICANT: DESAI, S.
; APPLICANT: DAILEY, S.
; TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.
; ZIP: 60065-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06965A
; FILING DATE: 19920821
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834PC.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-06965A-23

Query Match 100.0%; Score 40; DB 5; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 447 SLMAFTAAV 455

Search completed: August 23, 2002, 09:57:53
Job time: 203 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:49 ; Search time 78.82 seconds
(without alignments)
10.972 Million cell updates/sec

Title: US-08-854-825-34

Perfect score: 40

Sequence: 1 SLMAFTAAV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	492	2 PS0326	polyprotein - hepa
2	40	100.0	3011	1 GNMVC3	genome polyprotein
3	40	100.0	3011	1 GNMVCH	genome polyprotein
4	40	100.0	3011	1 S40770	genome polyprotein
5	36	90.0	3010	1 GNMVTC	genome polyprotein
6	36	90.0	3010	1 GNMVCJ	genome polyprotein
7	36	90.0	3010	1 A45573	genome polyprotein
8	36	90.0	3010	1 S18030	genome polyprotein
9	36	90.0	3010	1 GNMVTW	genome polyprotein
10	34	85.0	3014	1 JC5620	genome polyprotein
11	33	82.5	135	2 S67924	spore-wall fungal
12	33	82.5	402	2 B69843	conserved hypotet
13	33	82.5	471	2 Y78347	photosystem II chl
14	32	80.0	466	1 YTB5MA	methylerythrin A r
15	31	77.5	524	2 S35551	transcription fact
16	31	77.5	3033	1 JQ1303	genome polyprotein
17	31	77.5	3033	1 GNMVJ8	genome polyprotein
18	30	75.0	166	1 NVB02	vasopressin / neur
19	30	75.0	177	1 RKDWSU	ribulose-bisphosph
20	30	75.0	177	1 RKDWSB	ribulose-bisphosph
21	30	75.0	177	1 RKDWSA	ribulose-bisphosph
22	30	75.0	177	1 RKDWS4	ribulose-bisphosph
23	30	75.0	177	1 RKDWS6	ribulose-bisphosph
24	30	75.0	220	2 T21730	hypothetical prote
25	30	75.0	269	2 T26957	hypothetical prote
26	30	75.0	283	2 E88597	protein Y47D3B.6 l
27	30	75.0	310	2 T35754	probable branched
28	30	75.0	333	2 JX0343	triacylglycerol l1
29	30	75.0	375	2 A83802	NAD biosynthesis n

30 30 75.0 549 2 T15506 hypothetical prote
31 30 75.0 701 2 T28151 probable ABC-type
32 29 72.5 199 2 S57690 probable membrane
33 29 72.5 267 2 JN0845 enterohemolysin 1
34 29 72.5 324 2 S49586 cysteine synthase
35 29 72.5 333 2 S46984 indole-3-pyruvate
36 29 72.5 348 2 S52720 homeobox protein g
37 29 72.5 385 2 A72480 hypothetical prote
38 29 72.5 399 2 C95943 hypothetical prote
39 29 72.5 442 2 S50332 NADH dehydrogenase
40 29 72.5 484 2 S44739 C02C2.1 protein -
41 29 72.5 492 2 S09705 glucose transport
42 29 72.5 552 2 T51439 oligopeptide trans
43 29 72.5 588 2 E95951 conserved probable
44 29 72.5 802 2 T21315 hypothetical prote
45 29 72.5 902 2 E90270 conserved hypothet

ALIGNMENTS

RESULT 1

PS0326

polyprotein - hepatitis C virus (isolate Fla) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: PS0326

R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.

Gene 105, 167-172, 1991

A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genot

A:Reference number: PS0326; MUID:92039028

A:Accession: PS0326

A:Molecule type: genomic RNA

A:Residues: 1-492 <LIJ>

A:Cross-references: GB:M60220

A:Note: this sequence corresponds to nonstructural protein NS3 region

A:Note: translation of the nucleotide sequence is not complete

C:Superfamily: hepatitis C virus genome polyprotein

C:keywords: polyprotein

Query Match 100.0%; Score 40; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9

Db 340 SLMAFTAAV 348

RESULT 2

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (BC 3.4.21.98) (nonstru

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:DL0128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22

Query Match 100.0%; Score 40; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 1789 SLMAFTAAV 1797

RESULT 3
GNVTC

genome polypeptide - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A:Reference number: A41546; MUID:92052256

A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 23

Query Match 100.0%; Score 40; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 1789 SLMAFTAAV 1797

RESULT 4
S40770

genome polypeptide - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

A:Cross-references: EMBL:DL0749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116

A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 40; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
DB 1789 SLMAFTAAV 1797

RESULT 5
GNVTC

genome polypeptide - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEHX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>

A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPN>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2240,2242,2243,2244,2245,2246,2247,2248,2249,2250,2251,2252,2253,2254,2255,2256,2257,2258,2259,2260,2261,2262,2263,2264,2265,2266,2267,2268,2269,2270,2271,2272,2273,2274,2275,2276,2277,2278,2279,2280,2281,2282,2283,2284,2285,2286,2287,2288,2289,2290,2291,2292,2293,2294,2295,2296,2297,2298,2299,2300,2301,2302,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,2924,2925,2926,2927,2928,2929,2930,2931,2932,2933,2934,2935,2936,2937,2938,2939,2940,2941,2942,2943,2944,2945,2946,2947,2948,2949,2950,2951,2952,2953,2954,2955,2956,2957,2958,2959,

A:Residues: 1-547,'T','549-621','V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; envelope protein; nucleotide binding; P-loop; polyprotein; serin
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS

Query Match 90.0%; Score 36; DB 1; Length 3010;
Best Local Similarity 77.8%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
|||||||:
DB 1789 SLMAFTASI 1797

RESULT 9
GNWVFW
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 90.0%; Score 36; DB 1; Length 3010;
Best Local Similarity 77.8%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
|||||||:
DB 1789 SLMAFTASI 1797

RESULT 10
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593
A:Accession: JC5620
A:Molecule type: mRNA
A:Residues: 1-3014 <CHA>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:384-408/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitis C virus genome polyprotein
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 85.0%; Score 34; DB 1; Length 3014;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
|||||||:
DB 1790 TLMSTAAV 1798

RESULT 11
S67924
spore-wall fungal hydrophobin DewA - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997
C:Accession: S67924
R:Stringer, M.A.; Timberlake, W.E.
Mol. Microbiol. 16, 33-44, 1995
A:Title: dewA encodes a fungal hydrophobin component of the Aspergillus spore wall.
A:Reference number: S67924; MUID:95379492
A:Accession: S67924
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-135 <STR>
A:Cross-references: EMBL:U07935; NID:9533424; PID:9533425
C:Genetics:
A:Gene: dewA; CAN4
A:Map position: V
A:Introns: 97/1; 121/3

Query Match 82.5%; Score 33; DB 2; Length 135;
Best Local Similarity 87.5%; Pred. No. 5.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAA 8

```

Db      6 SLAFTAA 13
|||||
RESULT 12
B69843
Conserved hypothetical protein yjbb - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: B69843
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni,
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauerl
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: the complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: B69843
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-402 <KUN>
A:Cross-references: GB:299110; GB:AL009126; NID:g2633472; PIDN:CAB13005.1; PID:ell8168;
A:Experimental source: strain 168
A:Genetics:
A:Gene: yjbb

Query Match      82.5%; Score 33; DB 2; Length 402;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 LMAFTAAV 9
|||||
Db      235 LLAFTAAV 242

RESULT 13
S78347
Photosystem II chlorophyll a-binding protein psbc - Odontella sinensis chloroplast
N:Alternate names: photosystem II CP43 chlorophyll apoprotein
C:Species: chloroplast Odontella sinensis
C>Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 20-Jun-2000
C:Accession: S78347
R:Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A>Title: The Chloroplast Genome of a Chlorophyll a+c- containing Alga, Odontella sinensis
A:Reference number: S78347
A:Accession: S78347
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-471 <KOW>
A:Cross-references: EMBL:267753; NID:g1185127; PIDN:CAA91720.1; PID:g1185237
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C:Genetics:
A:Gene: psbc
A:Genome: chloroplast
C:Superfamily: photosystem II chlorophyll a-binding protein psbc
C:Keywords: chlorophyll; chloroplast; membrane-associated complex; phosphoprotein; photo
F:47-68/Domain: transmembrane #status predicted <TM1>
F:109-134/Domain: transmembrane #status predicted <TM2>
F:155-173/Domain: transmembrane #status predicted <TM3>
F:231-250/Domain: transmembrane #status predicted <TM4>
F:271-295/Domain: transmembrane #status predicted <TM5>
F:422-444/Domain: transmembrane #status predicted <TM6>

```

F:15/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.5%; Score 33; DB 2; Length 471;
 Best Local Similarity 77.8%; Pred. No. 18;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9

|||||
 Db 278 SLMGFTAAAL 286

RESULT 14

YTB5MA

methylenomycin A resistance protein mmr - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000

C:Accession: I40493; G69658; S2742

R:Putzer, H.; Gendron, N.; Grunberg-Manago, M.

EMBO J. 11, 3117-3127, 1992

A>Title: Co-ordinate expression of the two threonyl-tRNA synthetase genes in Bacillus

A:Reference number: I40493; MUID:92347349

A:Accession: I40493

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-466 <RES>

A:Cross-references: EMBL:X66121; NID:g40214; PIDN:CAA46908.1; PID:g40215

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033

A:Accession: G69658

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <KUN>

A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15784.1; PID:g26362

A:Experimental source: strain 168

C:Genetics:

A:Gene: mmr

C:Superfamily: tetracycline resistance protein

C:Keywords: antibiotic resistance; membrane protein

Query Match 80.0%; Score 32; DB 1; Length 466;

Best Local Similarity 87.5%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LMAFTAAV 9

|||||
 Db 236 LVAFTAAV 243

RESULT 15

S35551

transcription factor IIF chain RAP74 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 09-Dec-1993 #sequence_revision 26-May-1995 #text_change 21-Jul-2000

C:Accession: S35551

R:Gong, D.W.; Hasegawa, S.; Wada, K.; Roeder, R.G.; Nakatani, Y.; Horikoshi, M.

Nucleic Acids Res. 20, 6736, 1992

A>Title: Elucidation of three putative structural subdomains by comparison of primary

A;Reference number: S35551; MUID:93126122
 A;Accession: S35551
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-524 <GON>
 A;Cross-references: EMBL:Z17426; NID:965029; PIDN:CAA78999.1; PID:965030
 C;Genetics:
 A;Gene: RAP74
 C;Keywords: DNA binding; transcription regulation

Query Match 77.5%; Score 31; DB 2; Length 524;
 Best Local Similarity 87.5%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SLMAFTAA 8
 |||||
 Db 27 SLMAFNAA 34

Search completed: August 23, 2002, 09:56:51
 Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:24 ; Search time 37 Seconds
(without alignments)
9,418 Million cell updates/sec

Title: US-08-854-825-34
Perfect score: 40
Sequence: 1 SLMAFTAAV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	3011	1	POLG_HCV1
2	40	100.0	3011	1	POLG_HCVH
3	36	90.0	3010	1	POLG_HCVBK
4	36	90.0	3010	1	POLG_HCVJA
5	36	90.0	3010	1	POLG_HCVJT
6	36	90.0	3010	1	POLG_HCVTW
7	33	82.5	135	1	DEWA_EMENI
8	33	82.5	471	1	PSBC_ODOSI
9	32	80.0	466	1	MWR_BACSU
10	31	77.5	524	1	T2FA_XENLA
11	31	77.5	3033	1	POLG_HCVJ6
12	31	77.5	3033	1	POLG_HCVJ8
13	30	75.0	166	1	NEU2_BOVIN
14	30	75.0	177	1	RBS2_LENGI
15	30	75.0	177	1	RBS5_LENGI
16	30	75.0	177	1	RBS6_LENGI
17	30	75.0	375	1	ISCS_BACHD
18	29	72.5	199	1	YG4T_YEAST
19	29	72.5	348	1	GBX2_HUMAN
20	29	72.5	348	1	GBX2_MOUSE
21	29	72.5	484	1	YKHL_CAEEL
22	29	72.5	492	1	GTRI_MOUSE
23	28	70.0	77	1	VM21_YEAST
24	28	70.0	95	1	NCCV_ALCXH
25	28	70.0	176	1	DSBB_ECOLI
26	28	70.0	176	1	DSBB_SHIFL
27	28	70.0	276	1	BL01_ECOLI
28	28	70.0	307	1	BRAD_PSEAE
29	28	70.0	308	1	LIVH_ECOLI
30	28	70.0	308	1	LIVH_SALTY
31	28	70.0	330	1	GPR3_HUMAN
32	28	70.0	330	1	GPR3_MOUSE
33	28	70.0	337	1	NOD2_BRASN

RESULT 1

POLG_HCV1

ID POLG_HCV1

AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein c (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate 1) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus

OX NCBI_TaxID=11104;

RN [1]

RP MEDLINE=91172826; PubMed=1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Kuo G., Houghton M.;

RT "Genetic organization and diversity of the hepatitis C virus.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M62321; AAA45676.1; -;

DR PIR: A39166; GNRVC3.

DR HSP: P27958; JHEI.

DR MEROPS: S29.001; -;

DR MEROPS: U39.001; -;

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002531; HCV_NS1.

DR InterPro: IPR002518; HCV_NS2.

DR InterPro: IPR004109; HCV_NS3.

DR InterPro: IPR000745; HCV_NS4a.

P15979 gallus gall
P16733 human cytom
P31666 escherichia
Q01361 bos taurus
Q09449 caenorhabdi
P41647 pinus thunb
O9hvm1 pseudomonas
P40305 homo sapien
Q9d365 mus musculu
P56938 rhodobacter
P16967 canis fami
P31129 escherichia

34 28 70.0 345 1 HALF_CHICK
35 28 70.0 372 1 VGLM_HCMVA
36 28 70.0 409 1 YADE_ECOLI
37 28 70.0 487 1 CPDE_BOVIN
38 28 70.0 1551 1 YQ12_CAEEL
39 28 70.0 1756 1 YCF1_PINTH
40 27 67.5 91 1 RS20_PSEAE
41 27 67.5 122 1 INI7_HUMAN
42 27 67.5 180 1 SP22_MOUSE
43 27 67.5 284 1 COXX_RHOSH
44 27 67.5 286 1 SSRA_CANFA
45 27 67.5 296 1 YDEH_ECOLI

ALIGNMENTS

```

DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 323
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1635
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT CHAIN 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
FT CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
RESULT 2

```

```

POLG_HCVH
ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC !- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC !- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC !- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC !- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC !- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC !- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC !- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC !- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC !- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC !- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1A1V; 16-FEB-99.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -

```

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 747 809 PROTEIN P7.
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CB929CD94753 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9

|||||

Db 1789 SLMAFTAAV 1797

RESULT 3
POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P67); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers."
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetevy L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase."
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomsky J.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site."
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomel L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form."
RL Protein Sci. 7:837-847(1998).
CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M58335; AAA72945.1; -.
DR PIR; A38465; GNWVTC.
DR PDB; 1A1Q; 25-MAR-98.

DR PDB: 1JXP; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01539; HCV_core; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRp; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2768 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ CARBOHYD 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRG64;
 Query Match 90.0%; Score 36; DB 1; Length 3010;
 Best Local Similarity 77.8%; Pred. No. 8.3;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SLMAFTAAV 9
 DB 1789 SLMAFTASI 1797
 RESULT 4
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein E7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE Hepatitis C virus (isolate Japanese) (HCV)
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91089550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 [2]
 RN DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90208; BAAL4233.1; -.
 DR PIR; A39253; GNWVCJ.
 DR HSSP; P26663; LUXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR002166; HCV_NS4b.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01539; HCV_core; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRp; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.

DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4; 1.
DR Pfam; PF01001; HCV_NS4B; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELICG3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3010 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT SITE 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 90.08; Score 36; DB 1; Length 3010;
Best Local Similarity 77.8%; Pred. No. 8.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SLMAFTAAV 9
|||||
Db 1789 SLMAFTASI 1797

RESULT 5
POLG_HCVJT
ID POLG_HCVJT
AC Q00269; STANDARD; PRT; 3010 AA.
DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RL Virus Res. 23:39-53(1992)
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1',
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D11168; BAA01943.1; -
CC PIR; A45573; A45573.
CC HSSP; P26663; 1JXP.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRP.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00998; HCV_RdRP; 1.
CC Pfam; PF00271; helicase_C; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00492; HELICG3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
Transmembrane; Nonstructural protein; Hydroxylase; Serine protease.
INIT_MET 1 1
CHAIN 1 115
CHAIN 116 191
CHAIN 192 383
CHAIN 384 729
CHAIN 730 1006
CHAIN 1007 1615
CHAIN 1616 1862
CHAIN 1863 2013
CHAIN 2014 3010
TRANSMEM 347 369
ACT_SITE 1083 1083
ACT_SITE 1107 1107
ACT_SITE 1165 1165
NP_BIND 1230 1237
SITE 1316 1319
CARBOHYD 196 196
CARBOHYD 209 209
CARBOHYD 234 234
CARBOHYD 250 250
CARBOHYD 305 305
CARBOHYD 417 417
CARBOHYD 423 423
CARBOHYD 430 430
CARBOHYD 448 448
CARBOHYD 532 532
CARBOHYD 540 540
CARBOHYD 556 556
CARBOHYD 576 576
CARBOHYD 623 623
CARBOHYD 645 645
CARBOHYD 2041 2041
CARBOHYD 2077 2077
CARBOHYD 2240 2240
CARBOHYD 2529 2529
CARBOHYD 2788 2788
SEQUENCE 3010 AA; 326573 MW; 94A1C774350642BB CRC64;

Query Match 90.0%; Score 36; DB 1; Length 3010;
Best Local Similarity 77.8%; Pred. No. 8.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
|||||||
Db 1789 SLMAFTASI 1797

RESULT 6
POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT The Taiwanese hepatitis C virus genome: sequence determination and
mapping the 5' termini of viral genomic and antigenomic RNA.*;
RL Virology 188:102-113(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the p6
position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL: M84754; -; NOT ANNOTATED_CDS.
DR FIR; A40244; GNWVTW.
DR HSP; P26663; LJXP.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRp; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydroxylase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
FT SEQUENCE 3010 AA; 326573 MW; 94A1C774350642BB CRC64;

```

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 90.08; Score 36; DB 1; Length 3010;
Best Local Similarity 77.8%; Pred. No. 8.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTASI 1797

RESULT 7
DEWA_EMENI STANDARD; PRT; 135 AA.
AC P52750;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Spore-wall fungal hydrophobin DEWA precursor.
GN DEWA
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutriales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95379492; PubMed=7651135;
RA Stringer M.A., Timberlake W.E.;
RT "dewa encodes a fungal hydrophobin component of the Aspergillus spore wall."
RL Mol. Microbiol. 16:33-44(1995).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SPORE WALL.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES PREFERENTIALLY DURING ASEQUAL SPORULATION.
CC -1- SIMILARITY: BELONGS TO THE FUNGAL HYDROPHOBIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07935; AAC13762.1; -
DR InterPro; IPR001338; Hydrophobin.
DR Pfam; PF01185; Hydrophobin; 1.
DR SMART; SM00075; HYDRO; 1.
DR PROSITE; PS00956; HYDROPHOBIN; 1.
KW Cell wall; Glycoprotein; Signal.

```

```

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 135 SPORE-WALL FUNGAL HYDROPHOBIN DEWA.
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 135 AA; 13201 MW; F3B364748293A205 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 135;
Best Local Similarity 87.5%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAA 8
Db 6 SLMAFTAA 13

RESULT 8
PSBC_ODOSI STANDARD; PRT; 471 AA.
AC P49472;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CPA3).
GN PSBC.
OS Odontella sinensis.
OC Chloroplast.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Biddulphiophycidae; Eupodiscales; Eupodisaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga, Odontella sinensis."
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -1- FUNCTION: THE 43 kDa PROTEIN (P6) IS A COMPONENT OF THE CORE OF PHOTOSYSTEM II. IT IS A CHLOROPHYLL BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CHLOROPLAST THYLAKOID MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z67753; CAA91720.1; -
DR Mendel; 2635; ODOSI; psbC; 1.
DR InterPro; IPR000932; PSII.
DR Pfam; PF00421; PSII; 1.
KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Chloroplast;
KW Transmembrane.
SQ SEQUENCE 471 AA; 51878 MW; 340A424CE2AD2CFC CRC64;

Query Match 82.5%; Score 33; DB 1; Length 471;
Best Local Similarity 77.8%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 278 SLMGFTAA 286

RESULT 9
MMR_BACSU STANDARD; PRT; 466 AA.
ID MMR_BACSU
AC Q00538;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)

```



```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Methylenmycin A resistance protein (MMR peptide).
GN MMR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Putzer H., Gendron N., Grunberg-Manago M.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RESISTANCE TO THE EPOXIDE ANTIMBIOTIC METHYLENOMYCIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66121; CAA46908.1; -.
DR EMBL; 299123; CAB15784.1; -.
DR PIR; S22742; YTB5MA.
DR Subtilist; BG10420; mmmr.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR Antibiotic resistance; Transmembrane; Transport; Complete proteome.
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
SQ SEQUENCE 466 AA; 48845 MW; 7F77702100736799 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 466;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LMAFTAAV 9
   1:|||||
DB 236 LVAFTAAV 243

RESULT 10
T2FA_XENLA
ID T2FA_XENLA STANDARD; PRT; 524 AA.
AC Q04870;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcription initiation factor IIF, alpha subunit (TFIIF-alpha)
DE (Transcription initiation factor RAP74).
GN RAP74.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93126122; PubMed=1480494;
RA Gong D.-W., Hasegawa S., Wada K., Roeder R.G., Nakatani Y.,
RA Horikoshi M.;
RT "Elucidation of three putative structural subdomains by comparison of
RT primary structure of Xenopus and human RAP74.";
RL Nucleic Acids Res. 20:6736-6736(1992).
CC -!- FUNCTION: TFIIF IS A GENERAL TRANSCRIPTION INITIATION FACTOR THAT
CC BINDS TO RNA POLYMERASE II AND HELPS TO RECRUIT IT TO THE
CC INITIATION COMPLEX IN COLLABORATION WITH TFIIB. IT PROMOTES
CC TRANSCRIPTION ELONGATION.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: TO OTHER TRANSCRIPTION FACTOR IIF, ALPHA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z17426; CAA78999.1; -.
DR PIR; S35551; S35551.
DR TRANSFAC; T02171; -.
KW Transcription regulation; DNA-binding; Nuclear protein.
SQ SEQUENCE 524 AA; 58699 MW; 8CF3A74A3BF7BB0 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 524;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLMAFTAA 8
   1:|||||
DB 27 SLMAFNAA 34

RESULT 11
POLG_HCVJ6
ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier; comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.

```


CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D00944; BAA00792.1; -.
CC PIR: J01303; J01303.
CC HSP: P27958; IHEI.
CC MEROPS: S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; HCV_RdRP; 1.
CC Pfam: PF00271; helicase_C; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Envelope protein; Helicase; ATP-binding;
CC Core protein; Coat protein; Envelope protein; Hydrolase; Serine protease.
CC Transmembrane; Nonstructural
CC INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1230 1241
FT SITE 1324 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534

FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 77.5%; Score 31; DB 1; Length 3033;
Best local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1793 SMMAFSAAL 1801
1:111:111:

RESULT 12
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]
RP SEQUENCE FROM N.A.
RA Okamoto H., Tsuda K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D10988; BAA01761.1; -.
CC PIR: A40250; GNMVJ8.
CC HSP: P27958; IHEI.
CC MEROPS: S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro: IPR001410; DEAD.

DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5a.
 DR InterPro: IPR002522; HCV_RDRP.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RDRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 77.5%; Score 31; DB 1; Length 3033;
 Best Local Similarity 66.7%; Pred. No. 1.le+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SLMAFTAAV 9
 I:|||||:
 Db 1793 SMWFSAL 1801

RESULT 13
 ID NEU2_BOVIN STANDARD; PRT; 166 AA.
 AC P01180;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Vasopressin-neurophysin 2-copeptin precursor [Contains: Arg-
 DE vasopressin; Neurophysin 2 (Neurophysin-II); Copeptin].
 GN AVP.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87026235; PubMed=3768139;
 RA Rehbein M., Hillers M., Mohr E., Ivell R., Morley S., Schmale H.,
 RA Richter D.;
 RT "The neurohypophyseal hormones vasopressin and oxytocin. Precursor
 RT structure, synthesis and regulation.";
 RL Biol. Chem. Hoppe-Seyler 367:695-704(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84168167; PubMed=6709064;
 RA Ruppert S., Scherer G., Schutz G.;
 RT "Recent gene conversion involving bovine vasopressin and oxytocin
 RT precursor genes suggested by nucleotide sequence.";
 RL Nature 308:554-557(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82125432; PubMed=6276766;
 RA Land H., Schutz G., Schmale H., Richter D.;
 RT "Nucleotide sequence of cloned cDNA encoding bovine arginine
 RT vasopressin-neurophysin II precursor.";
 RL Nature 295:299-303(1982).
 RN [4]
 RP SEQUENCE OF 20-28.
 RA du Vigneaud V., Lawler H.C., Popenoe E.A.;
 RT "Enzymatic cleavage of glycineamide from vasopressin and a proposed
 RT structure for this pressor-antidiuretic hormone of the posterior
 RT pituitary.";
 RL J. Am. Chem. Soc. 75:4880-4881(1953).
 RN [5]
 RP SEQUENCE OF 20-28.
 RA Acher R., Chauvet J.;
 RT "The structure of bovine vasopressin.";
 RL Biochim. Biophys. Acta 12:487-488(1953).
 RN [6]
 RP SEQUENCE OF 32-126.
 RX MEDLINE=88076962; PubMed=3318825;
 RA Burman S., Breslow E., Chait B.T., Chaudhary T.;
 RT "Partial assignment of disulfide pairs in neurophysins.";
 RL Biochem. Biophys. Res. Commun. 148:827-833(1987).
 RN [7]
 RP SEQUENCE OF 32-126.
 RA Chauvet M.-T., Chauvet J., Acher R.;
 RT "The neurohypophyseal hormone-binding protein: complete amino-acid
 RT sequence of ovine and bovine MSEL-neurophysins.";
 RL Eur. J. Biochem. 69:475-485(1976).
 RN [8]
 RP SEQUENCE OF 32-126.
 RX MEDLINE=76135572; PubMed=1252249;
 RA Wu T.-C., Crumm S.E.;
 RT "Amino acid sequence of bovine neurophysin-II: a reinvestigation.";
 RL Biochem. Biophys. Res. Commun. 68:634-639(1976).
 RN [9]
 RP PRELIMINARY SEQUENCE (FETAL NEUROPHYSIN 2).
 RX MEDLINE=76118292; PubMed=1248642;
 RA Chauvet M.-T., Chauvet J., Acher R.;

RT "Foetal bovine MSEL-neurophysin: comparison with adult homologous
neurophysin.";
FEBS Lett. 62:89-92(1976).
[10]
RN DISULFIDE BONDS OF NEUROPHYSIN 2.
RX MEDLINE=73044843; PubMed=4564211;
RA Schlesinger D.H., Frangione B., Walter R.;
RT "Covalent structure of bovine neurophysin-II: localization of the
RT disulfide bonds";
Proc. Natl. Acad. Sci. U.S.A. 69:3350-3354(1972).
[11]
RN SEQUENCE OF 128-166
RX MEDLINE=79231438; PubMed=465021;
RA Smyth D.G., Massey D.E.;
RT "A new glycopeptide in pig, ox and sheep pituitary.";
RN Biochem. Biophys. Res. Commun. 87:1006-1010(1979).
[12]
RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF NEUROPHYSIN 2.
RX MEDLINE=91239543; PubMed=2034668;
RA Chen L.Q., Rose J.P., Breslow E., Yang D., Chang W.-R.,
RA Furey W.F. Jr., Sax M., Wang B.-C.;
RT "Crystal structure of a bovine neurophysin II dipeptide complex at
RT 2.8 A determined from the single-wavelength anomalous scattering
RT signal of an incorporated iodine atom";
Proc. Natl. Acad. Sci. U.S.A. 88:4240-4244(1991).
[13]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF NEUROPHYSIN 2.
RX MEDLINE=96163436; PubMed=8564543;
RA Rose J.P., Wu C.-K., Hsiao C.-D., Breslow E., Wang B.-C.;
RT "Crystal structure of the neurophysin-oxycytocin complex.";
RN Nat. Struct. Biol. 3:163-169(1996).
CC -!- FUNCTION: NEUROPHYSIN 2 SPECIFICALLY BINDS VASOPRESSIN.
CC KIDNEY. IT ALSO CAUSES VASOCONSTRICTION OF THE PERIPHERAL VESSELS.
CC -!- FUNCTION: VASOPRESSIN HAS A DIRECT ANTIDIURETIC ACTION ON THE
CC IN THE NEUROPHYSIN OF 7 TO 9 MONTH FETUSES & ITS SEQUENCE
CC APPEARS TO BE IDENTICAL WITH THAT OF THE ADULT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M25645; AAA30806.1; -;
DR EMBL; V00113; CAA23447.1; -;
DR EMBL; X00503; CAA25195.1; -;
DR PIR; A01442; NVBO2.
DR PIR; A29724; A29724.
DR PIR; S09580; S09580.
DR PDB; 2BN2; 16-FEB-99.
DR PDB; 1NFO; 12-FEB-97.
DR InterPro; IPR000981; Neurohypophys_horm.
DR Pfam; PF00220; hormone4; 1.
DR Pfam; PF00184; hormone5; 1.
DR PRINTS; PR00831; NEUROPHYSIN.
DR ProDom; PD001676; Neurohypophys_horm; 1.
DR SMART; SM00003; NH; 1.
DR PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
DR Hormone; Hypothalamus; Amidation; Cleavage on pair of basic residues;
KW Vasoconstrictor; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 19
FT PEPTIDE 20 28 ARG-VASOPRESSIN.
FT PEPTIDE 32 126
FT PEPTIDE 128 166 NEUROPHYSIN 2.
FT DISULFID 20 25
FT DISULFID 41 85
FT DISULFID 44 58
FT DISULFID 52 75
FT DISULFID 59 65

FT DISULFID 92 104
FT DISULFID 98 116
FT DISULFID 105 110
FT MOD_RES 28 28
FT CARBOHYD 133 133
FT VARIANT 120 120
SQ SEQUENCE 166 AA; 17325 MW; 0441DC255288DBF6 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 166;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAA 8
||:|:|:|:
Db 12 SLLAFTSA 19

RESULT 14
RBS2_LEMGI
ID RBS2_LEMGI STANDARD; PRT; 177 AA.
AC PI9308;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ribulose biphosphate carboxylase small chain SSU26, chloroplast
DE precursor (EC 4.1.1.39) (RuBisCO small subunit SSU26).
GN SSU26.
OS Lemna gibba (Swollen duckweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Lemna.
OX NCBI_TaxID=4470;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91355864; PubMed=2103442;
RA Silverthorne J., Wimpsee C.F., Yamada T., Rolfe S.A., Tobin E.M.;
RT "Differential expression of individual genes encoding the small
RT subunit of ribulose-1,5-bisphosphate carboxylase in Lemna gibba.";
RL Plant Mol. Biol. 15:49-58(1990).
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE. THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- MISCELLANEOUS: THIS PROTEIN IS CODED BY ONE MEMBER OF A SMALL
CC MULTIGENE FAMILY.
CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X17232; CAA35101.1; -;
DR PIR; S11680; RKDWS6.
DR HSSP; P00866; 4RUB.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR PRINTS; PR00152; RUBISCO SMALL.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
KW Multigene family.

```
FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 58 177 RIBULOSE BISPHTHOSPHATE CARBOXYLASE SMALL
FT CHAIN SSU26.
SQ SEQUENCE 177 AA; 19815 MW; 44D8151B354462B9 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 177;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1:11111111
4 SMMASTAAV 12

RESULT 15
RBS5_LEMGI
ID RBS5_LEMGI STANDARD; PRT; 177 AA.
AC P19311;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ribulose bisphtosphate carboxylase small chain SSU5A, chloroplast
DE precursor (EC 4.1.1.39) (RUBISCO small subunit SSU5A).
GN SSU5A.
OS Lemna gibba (Swollen duckweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Lemna.
OX NCBI_TaxID=4470;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91355864; PubMed=2103442;
RA Silverthorne J., Wimpee C.F., Yamada T., Rolfe S.A., Tobin E.M.;
RT "Differential expression of individual genes encoding the small
RT subunit of ribulose-1,5-bisphosphate carboxylase in Lemna gibba.";
RL Plant Mol. Biol. 15:49-58(1990).
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- MISCELLANEOUS: THIS PROTEIN IS CODED BY ONE MEMBER OF A SMALL
CC MULTIGENE FAMILY.
CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17230; CAA35099.1; -
DR PIR; S11683; RKWSA.
DR HSP; P00866; 4RUB.
DR InterPro; IPR000894; RUBISCO_small.
DR pfam; PF00101; RUBISCO_small; 1.
DR PRINTS; PR00152; RUBISCO SMALL.
DR ProDom; PD000290; RUBISCO_small; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
KW Multigene family.
FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 58 177 RIBULOSE BISPHTHOSPHATE CARBOXYLASE SMALL
FT CHAIN SSU5A.
```

```
SQ SEQUENCE 177 AA; 19802 MW; 7C6E35F3A1C5FCB7 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 177;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1:11111111
4 SMMASTAAV 12
```

Search completed: August 23, 2002, 09:55:25
Job time: 55 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:21:59 ; Search time 139.83 Seconds
(without alignments)
11.135 Million cell updates/sec

Title: US-08-854-825-34

Perfect score: 40

Sequence: 1 SLMAFTAAV 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	188	12	Q68578 hepatitis c
2	40	100.0	197	12	Q68586 hepatitis c
3	40	100.0	197	12	Q68579 hepatitis c
4	40	100.0	198	12	Q68577 hepatitis c
5	40	100.0	198	12	Q68580 hepatitis c
6	40	100.0	271	12	Q81573 hepatitis c
7	40	100.0	2436	12	Q81756 hepatitis c
8	40	100.0	3011	12	Q36579 hepatitis c
9	40	100.0	3011	12	Q36608 hepatitis c
10	40	100.0	3011	12	Q36609 hepatitis c
11	40	100.0	3011	12	Q36610 hepatitis c
12	40	100.0	3011	12	Q91FE5 hepatitis c
13	40	100.0	3011	12	Q9ELS8 hepatitis c
14	40	100.0	3011	12	Q9DIT6 hepatitis c
15	40	100.0	3011	12	Q03463 hepatitis c
16	40	100.0	3015	12	Q9PWX5 hepatitis c

17	40	100.0	3015	12	Q9PWU9 hepatitis c
18	39	97.5	191	12	Q68584 hepatitis c
19	37	92.5	195	12	Q68582 hepatitis c
20	37	92.5	198	12	Q68583 hepatitis c
21	37	92.5	198	12	Q68585 hepatitis c
22	37	92.5	351	2	Q9LY77 pseudomonas
23	37	92.5	425	12	Q68344 hepatitis c
24	37	92.5	3008	12	Q39929 hepatitis c
25	37	92.5	3010	12	Q9DPE6 hepatitis c
26	37	92.5	3010	12	Q9DPTD9 hepatitis c
27	37	92.5	3010	12	Q9QIX2 hepatitis c
28	37	92.5	3010	12	Q9QIX1 hepatitis c
29	37	92.5	3010	12	Q9J3I0 hepatitis c
30	37	92.5	3010	12	Q9J3H9 hepatitis c
31	37	92.5	3010	12	Q9J3H0 hepatitis c
32	37	92.5	3010	12	Q9J3G4 hepatitis c
33	37	92.5	3010	12	Q81760 hepatitis c
34	37	92.5	3011	12	Q81754 hepatitis c
35	37	92.5	3011	12	Q9WIK7 hepatitis c
36	37	92.5	3012	12	Q9WIK7 hepatitis c
37	37	92.5	3019	12	Q68801 hepatitis c
38	37	92.5	3021	12	Q92933 hepatitis c
39	37	92.5	3021	12	Q81258 hepatitis c
40	37	92.5	3021	12	Q81495 hepatitis c
41	37	92.5	3021	12	Q68870 hepatitis c
42	37	92.5	3023	12	Q81487 hepatitis c
43	36	90.0	1186	12	Q81755 hepatitis c
44	36	90.0	1805	12	Q41809 hepatitis c
45	35	90.0	2284	12	Q81817 hepatitis c

ALIGNMENTS

RESULT 1

Q68578	PRELIMINARY;	PRT;	188 AA.
ID	Q68578		
AC	Q68578;		
DT	01-NOV-1996 (TRENBLrel. 01, Created)		
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)		
DE	NS4 PROTEIN (FRAGMENT).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BB136;		
RA	Songsivilai S., Leelawiat W., Dharakul T.;		
RT	"Molecular cloning and expression of the NS4 gene from the Thai isolates of hepatitis C virus."		
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; U55287; AB00218.1;		
DR	InterPro; IPR000745; HCV_NS4a.		
DR	InterPro; IPR001490; HCV_NS4b.		
DR	Pfam; PF01006; HCV_NS4a; 1.		
DR	Pfam; PF01001; HCV_NS4b; 1.		
FT	NON_TER		
FT	NON_TER		
SQ	SEQUENCE	188 AA; 19033 MW; CE63B2244FB31C20 CRC64;	

Query Match 100.0%; Score 40; DB 12; Length 188;

Best Local Similarity 100.0%; Pred. No. 0.43; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy 1 SLMAFTAAV 9

Db 26 SLMAFTAAV 34
|||||||

RESULT 2

```

Q68586
ID Q68586 PRELIMINARY; PRT; 197 AA.
AC Q68586;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB143;
RA Songsivilai S., Leelawiat W., Dharakul T.;
RT "Molecular cloning and expression of the NS4 gene from the Thai
RT isolates of hepatitis C virus.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55295; AAB00226.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
FT NON_TER 1
FT NON_TER 197
FT NON_TER 197
SQ SEQUENCE 197 AA; 20086 MW; CF8EA3DEDB7B97B5 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 27 SLMAFTAAV 35

RESULT 3
Q68579
ID Q68579 PRELIMINARY; PRT; 197 AA.
AC Q68579;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB143;
RA Songsivilai S., Leelawiat W., Dharakul T.;
RT "Molecular cloning and expression of the NS4 gene from the Thai
RT isolates of hepatitis C virus.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55288; AAB00219.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
FT NON_TER 1
FT NON_TER 197
FT NON_TER 197
SQ SEQUENCE 197 AA; 19908 MW; BA4D248A2F0AA9D6 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 29 SLMAFTAAV 37

RESULT 4
Q68577
ID Q68577 PRELIMINARY; PRT; 198 AA.
AC Q68577;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB10;
RA Songsivilai S., Leelawiat W., Dharakul T.;
RT "Molecular cloning and expression of the NS4 gene from the Thai
RT isolates of hepatitis C virus.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55286; AAB00217.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
FT NON_TER 1
FT NON_TER 198
FT NON_TER 198
SQ SEQUENCE 198 AA; 19984 MW; 2F5C04B9D326BE55 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 198;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 30 SLMAFTAAV 38

RESULT 5
Q68580
ID Q68580 PRELIMINARY; PRT; 198 AA.
AC Q68580;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB150;
RA Songsivilai S., Leelawiat W., Dharakul T.;
RT "Molecular cloning and expression of the NS4 gene from the Thai
RT isolates of hepatitis C virus.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55289; AAB00220.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
FT NON_TER 1
FT NON_TER 198
FT NON_TER 198
SQ SEQUENCE 198 AA; 19908 MW; 80D07F5A9168B67E CRC64;

Query Match 100.0%; Score 40; DB 12; Length 198;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 28 SLMAFTAAV 36

```



```
RESULT 6
Q81573 ID Q81573 PRELIMINARY; PRT; 271 AA.
AC Q81573;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NONSTRUCTURAL PROTEIN (FRAGMENT).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD2;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein."
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86766; AAA45616.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
SQ SEQUENCE 271 AA; 29796 MW; 72BEC54E6877CD4 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 271;
Best Local Similarity 100.0%; Pred. No. 0.63; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 SLMAFTAAV 9
Db 262 SLMAFTAAV 270
|||||

RESULT 7
Q81756 ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)] (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC Choo Q.-L., Richman K., Han J.;
RA "The nucleotide sequence of the Hepatitis C viral genome.";
RT Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; 1AIV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
```

```
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RDRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR PRINTS; PR00284; TOXIN.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 2436;
Best Local Similarity 100.0%; Pred. No. 5.6; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 SLMAFTAAV 9
Db 1339 SLMAFTAAV 1347
|||||

RESULT 8
O36579 ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)]
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA."
RL Science 277:570-574(1997).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF009606; AAB66324.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RDRP; 1.
```

DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C53C1B9 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
|||||

RESULT 9
Q36608 PRELIMINARY; PRT; 3011 AA.
ID O36608;
AC O36608;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)].
DE (NS1)].
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
RL -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
CC EMBL; AF011751; AAB67036.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
|||||

RESULT 9
Q36608 PRELIMINARY; PRT; 3011 AA.
ID O36608;
AC O36608;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)].
DE (NS1)].
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
RL -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
CC EMBL; AF011751; AAB67036.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
|||||

RESULT 9
Q36608 PRELIMINARY; PRT; 3011 AA.
ID O36608;
AC O36608;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)].
DE (NS1)].
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
RL -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
CC EMBL; AF011751; AAB67037.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327261 MW; 10D1C9702CA9B5DC CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
|||||

```
Db 1789 SLMAFTAAV 1797
RESULT 11
O36610 PRELIMINARY; PRT; 3011 AA.
AC O36610;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF011753; AAB67038.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002538; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327220 MW; 293F91430A0D4067 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
RESULT 12
Q91FE5 PRELIMINARY; PRT; 3011 AA.
ID Q91FE5;
AC Q91FE5;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLONEL;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
RESULT 13
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
ID Q9ELS8;
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLONEL;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLMAFTAAV 9
Db 1789 SLMAFTAAV 1797
RESULT 14
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
ID Q9ELS8;
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLONEL;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;
```

DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR002166; HCV_RdRP.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDc; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
| | | | | | | | | |
Db 1789 SLMAFTAAV 1797

RESULT 14
Q9DIT6 PRELIMINARY; PRT; 3011 AA.
AC Q9DIT6;
DT 01-WAR-2001 (TREMBLrel. 16, Created)
DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)].
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RY Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; D10749; BAA01582.1; -.
DR HSP; P27958; IHEI.
DR MEROPS; S29.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.

Query Match 100.0%; Score 40; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
| | | | | | | | | |
Db 1789 SLMAFTAAV 1797

RESULT 15
Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)].
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RY Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; D10749; BAA01582.1; -.
DR HSP; P27958; IHEI.
DR MEROPS; S29.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RdRP; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
 KW Transmembrane.
 SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 100.0%; Score 40; DB 12; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLMAFTAAV 9
 DB 1789 SLMAFTAAV 1797

Search completed: August 23, 2002, 10:22:01
 Job time: 1501 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:54:30 ; Search time 179.72 Seconds
(without alignments)
6.180 Million cell updates/sec

Title: US-08-854-825-1
Perfect score: 51
Sequence: 1 ADLMGYIPLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	100.0	10	16 AAR84570	Cytotoxic T-cell e
2	51	100.0	10	18 AAW39445	Human HCV1core-131
3	51	100.0	10	20 AAY10135	T cell epitope/MHC
4	51	100.0	11	22 AAJ00276	Hepatitis C virus
5	51	100.0	11	22 AAJ01550	Hepatitis C virus
6	51	100.0	15	19 AAW85270	Helper T-cell pept
7	51	100.0	15	22 AAJ03147	Hepatitis C virus
8	51	100.0	15	22 AAJ03178	Hepatitis C virus
9	51	100.0	15	22 AAJ03485	Hepatitis C virus
10	51	100.0	15	22 AAJ03985	Hepatitis C virus
11	51	100.0	16	16 AAR87651	Hepatitis C virus

12	51	100.0	16	20 AAY25210	HCV viral core pro
13	51	100.0	20	15 AAR56610	HCV peptide NP-13.
14	51	100.0	20	15 AAR56611	HCV peptide NP-14.
15	51	100.0	20	16 AAR84500	Hepatitis C virus
16	51	100.0	20	22 AAR82051	HCV antigen, core
17	51	100.0	135	21 AAB18534	Protein encoded by
18	51	100.0	150	16 AAR81946	Hepatitis C virus
19	51	100.0	150	18 AAW33688	Hepatitis C virus
20	51	100.0	150	20 AAW70830	Hepatitis C virus
21	51	100.0	150	20 AAW80518	Hepatitis C virus
22	51	100.0	154	12 AAR11274	Hepatitis C virus
23	51	100.0	154	16 AAR66615	Hepatitis C virus
24	51	100.0	154	16 AAR66620	Hepatitis C virus
25	51	100.0	154	20 AAY14970	HCV J7 C/E domain
26	51	100.0	154	20 AAY01614	Protein encoded by
27	51	100.0	154	20 AAY01609	Protein encoded by
28	51	100.0	154	20 AAW30583	Hepatitis C virus
29	51	100.0	158	17 AAR96558	Hepatitis C virus
30	51	100.0	161	19 AAW62658	Hepatitis C virus
31	51	100.0	166	15 AAR63374	Hepatitis C virus
32	51	100.0	166	15 AAR63359	Hepatitis C virus
33	51	100.0	166	15 AAR63353	Hepatitis C virus
34	51	100.0	169	15 AAR63291	Polypeptide encode
35	51	100.0	169	15 AAR63292	Polypeptide encode
36	51	100.0	169	21 AAB18535	Protein encoded by
37	51	100.0	178	17 AAR92947	Hepatitis C virus
38	51	100.0	178	19 AAW62823	HCV core protein p
39	51	100.0	182	19 AAW62659	HCV core protein (
40	51	100.0	190	13 AAR25123	Hepatitis C virus
41	51	100.0	190	17 AAR92948	Hepatitis C virus
42	51	100.0	191	13 AAR24525	HCV in plasmid pKM
43	51	100.0	191	14 AAW41754	Hepatitis C virus
44	51	100.0	191	14 AAR35758	Core protein regio
45	51	100.0	191	14 AAR44010	Hepatitis C Virus

ALIGNMENTS

RESULT 1	
ID AAR84570	standard; peptide; 10 AA.
AC AAR84570;	
XX	
DT 25-APR-1996	(first entry)
XX	
DE Cytotoxic T-cell epitope, aa 131-140 of HCV-1 core region.	
XX	
KW Hepatitis C virus; HCV-1; epitope; vaccine; immunogen.	
XX	
OS Hepatitis C virus.	
XX	
PN WO9525122-A1.	
XX	
PD 21-SEP-1995.	
XX	
PF 16-MAR-1995;	95WO-US03224.
XX	
PR 17-MAR-1994;	94US-0214650.
XX	
PA (SCRI) SCRIPPS RES INST.	
XX	
PI Cerny A, Chisari FV;	
XX	
DR WPI; 1995-336941/43.	
XX	
PT Novel molecule comprising a cytotoxic T cell epitope - used to	
PT vaccinate against hepatitis C viral infection	
XX	
PS Claim 1; Page 60; 85pp; English.	
XX	
CC AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides	

CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
 CC were tested for peptide specific cytotoxic T-cell activity. The
 CC peptides AAW84570-77 were found to have substantial homology with a T-
 CC cell epitope and are useful in vaccines against HCV infection.

XX Sequence 10 AA;

Query Match 100.0%; Score 51; DB 16; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00032; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
 DB | | | | | | | | | |

RESULT 2

AAW39445 ID AAW39445 standard; peptide; 10 AA.

XX AC AAW39445;

DT 11-JUN-1998 (first entry)

XX Human HCVlc core-131 immunogenic peptide.

XX T cell epitope; immune response; human leukocyte antigen; HLA Class I;
 KW vaccine; immunogenic; major histocompatibility complex; MHC; B cell;
 KW disease; anti-tumour; anti-viral.

XX Synthetic.

OS Homo sapiens.

XX WO9741440-A1.

XX 06-NOV-1997.

XX 28-APR-1997; 97WO-NL00229.

XX 23-DEC-1996; 96EP-0203670.

XX 26-APR-1996; 96EP-0201145.

XX (UYLE-) RIJKSUNIV LEIDEN.

PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.

XX Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;

XX WPI; 1997-549891/50.

XX Method of selecting T cell peptide epitope(s) - by measuring the
 PT stability of HLA class I-peptide complexes on intact B cells

XX Example 2; Page 27; 109pp; English.

XX Peptides AAW39430-W39734 are used in a novel method for the selection of
 CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
 CC method involves the identification of peptide sequences capable of
 CC binding to an HLA (human leukocyte antigen) class I molecule and
 CC measuring the binding of this epitope peptide to the HLA class I
 CC peptide. The stability of binding of the peptide and MHC (major
 CC histocompatibility complex) class I molecule is measured on intact human
 CC B cells carrying the MHC molecule at their cell surfaces. The method can
 CC be used to select peptide epitopes for generating vaccines against a
 CC disease associated with the polypeptide, e.g. cancers or AIDS. The
 CC peptide epitopes are especially T-cell peptide epitopes with strong
 CC anti-tumour and anti-viral immune responses. Peptide AAW39445 is the
 CC HCVlc core-131 peptide which is a HLA-A*0201 restricted cytotoxic T
 CC lymphocyte (CTL) epitope.

XX Sequence 10 AA;

Query Match 100.0%; Score 51; DB 18; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00032; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
 DB | | | | | | | | | |

RESULT 3

AAAY10135 ID AAY10135 standard; Peptide; 10 AA.

XX AC AAY10135;

XX 12-MAY-1999 (first entry)

XX T cell epitope/MHC ligand SEQ ID NO:65.

XX Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX Synthetic.

OS Hepatitis C virus.

XX WO9902183-A2.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-US14289.

XX 10-DEC-1997; 97US-0988320.

XX 10-JUL-1997; 97CA-2209815.

XX (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX Kuendig TM, Simard JUL;

XX WPI; 1999-120514/10.

XX Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX Disclosure; Page 25; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX Sequence 10 AA;

Query Match 100.0%; Score 51; DB 20; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00032; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
 DB | | | | | | | | | |


```
Db      1 adlmgyp1v 10

RESULT  4
AAJ00276
ID      AAJ00276 standard; Peptide; 11 AA.
XX
AC      AAJ00276;
XX
DT      02-JUL-2001 (first entry)
XX
DE      Hepatitis C virus epitope #267.
XX
KW      Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW      antiviral.
XX
OS      Hepatitis C virus.
XX
PN      WO200121189-A1.
XX
PD      29-MAR-2001.
XX
PF      19-JUL-2000; 2000WO-US19774.
XX
PR      19-JUL-1999; 99US-0357737.
XX
PA      (EPIM-) EPIMMUNE INC.
XX
PI      Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI      Baker DM, Celis E, Kubo RT, Grey HM;
XX
WPI:    2001-308046/32.
XX
PT      A new composition useful as a vaccines against hepatitis C virus -
PS      Disclosure; Page 139; 214pp; English.
XX
CC      The present invention describes a composition comprising a prepared
CC      hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC      These are derived from HCV HLA-binding motifs. They are useful in
CC      vaccines for the prevention and treatment of HCV infection in humans. The
CC      present sequence is an epitope used in the disclosure of the invention.
XX
SQ      Sequence 11 AA;

      Query Match      100.0%; Score 51; DB 22; Length 11;
      Best Local Similarity 100.0%; Pred. No. 0.00036;
      Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ADLMGYPLV 10
      |||||
Db      2 adlmgyp1v 11

RESULT  6
AAW85270
ID      AAW85270 standard; peptide; 15 AA.
XX
AC      AAW85270;
XX
DT      16-FEB-1999 (first entry)
XX
DE      Helper T-cell peptide derived from a Hepatitis C virus protein.
XX
KW      Helper T-cell peptide; human leucocyte antigen; HLA; DR4*4; DR1; DR7;
KW      cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;
KW      acquired immune deficiency syndrome; malaria; cancer;
KW      allograft rejection; allergy; Lyme disease; hepatitis;
KW      post-streptococcal endocarditis; glomerulonephritis;
KW      food hypersensitivity.
XX
OS      Synthetic.
OS      Hepatitis C virus.
XX
PN      WO9832456-A1.
XX
PD      30-JUL-1998.
XX
PF      23-JAN-1998; 98WO-US01373.
XX
PR      07-FEB-1997; 97US-0037432.
PR      23-JAN-1997; 97US-0036713.
XX
PA      (EPIM-) EPIMMUNE INC.
XX
PI      Sette A, Sidney J, Southwood S;
XX
WPI:    1998-427679/36.
XX
PT      Composition containing peptide that induces cytotoxic T lymphocyte
PT      response, and helper peptide - can bind to human leucocyte antigen
PT      alleles, used to treat or prevent cancers, parasitic infections and
PT      autoimmune disease

      Query Match      100.0%; Score 51; DB 22; Length 11;
      Best Local Similarity 100.0%; Pred. No. 0.00036;
      Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ADLMGYPLV 10
      |||||
Db      2 adlmgyp1v 11

RESULT  5
AAJ01550
ID      AAJ01550 standard; Peptide; 11 AA.
XX
AC      AAJ01550;
XX
DT      02-JUL-2001 (first entry)
XX
DE      Hepatitis C virus epitope #1541.
XX
KW      Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW      antiviral.
XX
OS      Hepatitis C virus.
XX
PN      WO200121189-A1.
XX
PD      29-MAR-2001.
```

XX Claim 11: Page 39; 51pp; English.

XX AAW85138-283 represent helper T-cell peptides, which can bind to the

CC human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides

CC are used in the course of the invention. The specification describes

CC peptides that induce a cytotoxic T lymphocyte (CTL) response, and

CC T-helper peptides, that are used together to generate a CTL response for

CC the treatment or prevention of viral, fungal, bacterial or parasitic

CC infections (e.g. hepatitis, acquired immune deficiency syndrome or

CC malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate

CC cancer or condyloma acuminatum). Helper T-cell peptides may be used

CC alone to induce a helper T cell response, e.g. in cases of autoimmune

CC disease, allograft rejection, allergy, Lyme disease, hepatitis,

CC post-streptococcal endocarditis, glomerulonephritis and food

CC hypersensitivity.

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 51; DB 19; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.00052;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

DB | | | | | | | |

1 adlmgyiplv 10

RESULT 7

AAJ03147

ID AAJ03147 standard; Peptide; 15 AA.

XX

AC AAJ03147;

XX

DT 02-JUL-2001 (first entry)

XX

DE Hepatitis C virus epitope #3138.

XX

XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;

KW antiviral.

XX

OS Hepatitis C virus.

XX

XX Hepatitis C virus.

XX

PN WO200121189-A1.

XX

PD 29-MAR-2001.

XX

XX 19-JUL-2000; 2000WO-US19774.

PF

XX 19-JUL-1999; 99US-0357737.

PR

XX (EPIM-) EPIMMUNE INC.

PA

XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;

PI Baker DM, Cellis E, Kubo RT, Grey HM;

XX

XX WPI; 2001-308046/32.

DR

XX A new composition useful as a vaccines against hepatitis C virus -

PT

XX Disclosure; Page 175; 214pp; English.

XX

CC The present invention describes a composition comprising a prepared

CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.

CC These are derived from HCV HLA-binding motifs. They are useful in

CC vaccines for the prevention and treatment of HCV infection in humans. The

CC present sequence is an epitope used in the disclosure of the invention.

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 51; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.00052;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

DB | | | | | | | |

1 adlmgyiplv 10

RESULT 9

AAJ03485

ID AAJ03485 standard; Peptide; 15 AA.

XX

AC AAJ03485;

XX

DT 02-JUL-2001 (first entry)

XX

DE Hepatitis C virus epitope #3476.

XX

XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;

KW antiviral.

XX

Best Local Similarity 100.0%; Pred. No. 0.00052;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

DB | | | | | | | |

2 adlmgyiplv 11

RESULT 8

AAJ03178

ID AAJ03178 standard; Peptide; 15 AA.

XX

AC AAJ03178;

XX

DT 02-JUL-2001 (first entry)

XX

DE Hepatitis C virus epitope #3169.

XX

XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;

KW antiviral.

XX

OS Hepatitis C virus.

XX

XX WO200121189-A1.

PN

XX 29-MAR-2001.

PD

XX 19-JUL-2000; 2000WO-US19774.

PF

XX 19-JUL-1999; 99US-0357737.

PR

XX (EPIM-) EPIMMUNE INC.

PA

XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;

PI Baker DM, Cellis E, Kubo RT, Grey HM;

XX

XX WPI; 2001-308046/32.

DR

XX A new composition useful as a vaccines against hepatitis C virus -

PT

XX Disclosure; Page 175; 214pp; English.

XX

CC The present invention describes a composition comprising a prepared

CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.

CC These are derived from HCV HLA-binding motifs. They are useful in

CC vaccines for the prevention and treatment of HCV infection in humans. The

CC present sequence is an epitope used in the disclosure of the invention.

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 51; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.00052;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

DB | | | | | | | |

1 adlmgyiplv 10

RESULT 9

AAJ03485

ID AAJ03485 standard; Peptide; 15 AA.

XX

AC AAJ03485;

XX

DT 02-JUL-2001 (first entry)

XX

DE Hepatitis C virus epitope #3476.

XX

XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;

KW antiviral.

XX

```
OS Hepatitis C virus.
XX WO200121189-A1.
XX 29-MAR-2001.
XX 19-JUL-2000; 2000WO-US19774.
XX 19-JUL-1999; 99US-0357737.
XX (EPIM-) EPIMMUNE INC.
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX WPI; 2001-308046/32.
XX A new composition useful as a vaccines against hepatitis C virus -
PT Disclosure; Page 178; 214pp; English.
XX
XX The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 15 AA;
SQ

Query Match 100.0%; Score 51; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ADLMGYIPLV 10
Db 1 adlmgyiplv 10
|||||
1 adlmgyiplv 10

RESULT 10
AAJ03985
ID AAJ03985 standard; Peptide; 15 AA.
XX
XX AAJ03985;
XX
XX 02-JUL-2001 (first entry)
XX
XX Hepatitis C virus epitope #3976.
XX
XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
XX Hepatitis C virus.
OS
XX WO200121189-A1.
XX
XX 29-MAR-2001.
XX
XX 19-JUL-2000; 2000WO-US19774.
XX
XX 19-JUL-1999; 99US-0357737.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX WPI; 2001-308046/32.
XX
XX A new composition useful as a vaccines against hepatitis C virus -
PT Example 5; Page 198; 214pp; English.
XX
```

```
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 15 AA;
SQ

Query Match 100.0%; Score 51; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ADLMGYIPLV 10
Db 1 adlmgyiplv 10
|||||
1 adlmgyiplv 10

RESULT 11
AAR87651
ID AAR87651 standard; peptide; 16 AA.
XX
XX AAR87651;
XX
XX 20-MAY-1996 (first entry)
XX
XX Hepatitis C virus core peptide used for treating HCV infection.
DE
XX Hepatitis C virus; cytotoxic T lymphocyte; vaccine; immunisation;
KW diagnosis; HCV; CTL; core; HLA.
XX
XX Hepatitis C virus.
OS
XX WO9527733-A1.
XX
XX 19-OCT-1995.
XX
XX 07-APR-1995; 95WO-US03935.
XX
XX 08-APR-1994; 94US-0224973.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Berzofsky JA, Feinstein S, Shirai M;
PI WPI; 1995-366355/47.
XX
XX Hepatitis C virus core peptide(s) stimulate cytotoxic T lymphocyte
PT response - used for prevention, treatment or diagnosis of HCV
PT infection
XX
XX Claim 5; Page 45; 58pp; English.
PS
XX
XX AAR87651-53 are hepatitis C virus (HCV) core region peptides which can
CC be used for the immunisation of an individual against HCV. They can
CC also be used for diagnosing exposure of a patient to HCV or for
CC predicting a patient's clinical course following HCV infection. The
CC peptides provoke the activation of cytotoxic T cells which attack
CC any cell infected with HCV.
XX
XX Sequence 16 AA;
SQ

Query Match 100.0%; Score 51; DB 16; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00056;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ADLMGYIPLV 10
Db 3 adlmgyiplv 12
|||||
3 adlmgyiplv 12

RESULT 12
```

AAV25210
 ID AAY25210 standard; peptide; 16 AA.
 XX
 AC AAY25210;
 XX
 DT 03-SEP-1999 (first entry)
 XX
 DE HCV viral core protein peptide fragment 1.
 XX
 KW Heat shock protein; HSP; complex; denatured protein matrix; antigen;
 KW vaccine; allergic disease; treatment; susceptibility; Th2; skin rash;
 KW allergic reaction; asthma; viral core protein.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9929182-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 04-DEC-1998; 98WO-US25734.
 XX
 PR 05-DEC-1997; 97US-0986234.
 PR 05-DEC-1997; 97US-0985548.
 XX
 PA (UYNE-) UNIV NEW MEXICO STATE.
 XX
 PI Moseley PL, Wallen ES;
 XX
 DR WPI; 1999-394912/33.
 XX
 PT Synthesizing heat shock protein complexes using a denatured protein
 PT matrix
 XX
 PS Example 1; Fig 1B; 33pp; English.
 XX
 CC This invention describes a novel method for synthesizing heat shock
 CC protein (HSP) complexes comprising adding a heat shock protein to a
 CC denatured protein matrix for binding, and adding a complexing solution
 CC comprising a peptide to elute a heat shock protein-peptide complex. A
 CC HSP-antigen complex is useful as a vaccine for treating an allergic
 CC disease (in a mammal, preferably a human) to reduce susceptibility of
 CC the Th2 response, the complex comprising a HSP-antigenic peptide complex.
 CC The complex is administered to prevent a mammal from having an allergic
 CC reaction to an allergic disease, or administered to a mammal having an
 CC allergic disease, to reduce the allergic reactions. Allergic diseases
 CC include asthma and skin rashes. Prior art methods or preventing/treating
 CC allergic diseases include antihistamines which treat only the symptoms,
 CC corticosteroids which have severe side effects and desensitization
 CC therapy which has limited uses. The new method also allows more
 CC flexibility of use of peptide-based vaccines, as prior art HSP-based
 CC vaccines require isolation from a portion of the tumour itself. This
 CC sequence represents a peptide fragment from the HCV viral core protein
 CC which is used in the method of the invention.
 XX
 SQ Sequence 16 AA;
 XX
 Query Match 100.0%; Score 51; DB 20; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.00056;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ADLMGYIPLV 10
 DB 3 adlmgyiplv 12
 |||||||||
 RESULT 13
 AAR56610
 ID AAR56610 standard; protein; 20 AA.
 XX
 AC AAR56610;
 XX
 DT 24-MAR-1995 (first entry)
 XX

XX HCV peptide NP-13.
 DE
 XX Hepatitis C virus; HCV; core protein; epitope; T-cell; immunity;
 KW CD8; CD4.
 KW
 XX Hepatitis C virus.
 OS
 XX JP06199894-A.
 PN
 PD 19-JUL-1994.
 XX
 PF 02-APR-1993; 93JP-0076791.
 XX
 PR 27-AUG-1992; 92JP-0228965.
 PR 10-NOV-1992; 92JP-0299691.
 XX
 PA (ASAH) ASAH KASEI KOGYO KK.
 PA (INOM/) INOMAWARI M.
 XX
 DR WPI; 1994-269451/33.
 XX
 PT T cell epitope present in the core protein region of Hepatitis C
 PT virus (HCV) - used for activation of cellular immunity mechanisms
 XX
 PS Disclosure; Fig 2; 14pp; Japanese.
 XX
 CC A T-cell stimulating peptide is claimed which is ca. 5-20 amino
 CC acids long and part of the core protein of HCV and is recognised by
 CC and stimulates T-cells. A CD8-positive T-cell stimulating peptide
 CC has the sequence given in AAR56606 (NP-9) or AAR56616 (9MA). A CD4-
 CC positive T-cell stimulating peptide has the sequence given in AAR56609
 CC (NP-12) or AAR56614 (NP-17). Synthetic peptide mixts. (Mix A: NP-1 -
 CC NP-5, Mix B: NP-6 - NP-10, Mix C: NP-11 - NP14, MixD: NP-15 - NP-18)
 CC were used in experiments to evaluate activity.
 CC Peptides 9MA, 9CM and 9CA (AAR56616-18) are variants of peptide NP-9.
 XX
 SQ Sequence 20 AA;
 XX
 Query Match 100.0%; Score 51; DB 15; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00073;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ADLMGYIPLV 10
 DB 11 adlmgyiplv 20
 |||||||||
 RESULT 14
 AAR56611
 ID AAR56611 standard; protein; 20 AA.
 XX
 AC AAR56611;
 XX
 DT 24-MAR-1995 (first entry)
 XX
 DE HCV peptide NP-14.
 XX
 KW Hepatitis C virus; HCV; core protein; epitope; T-cell; immunity;
 KW CD8; CD4.
 KW
 XX Hepatitis C virus.
 OS
 XX JP06199894-A.
 PN
 PD 19-JUL-1994.
 XX
 PF 02-APR-1993; 93JP-0076791.
 XX
 PR 27-AUG-1992; 92JP-0228965.
 PR 10-NOV-1992; 92JP-0299691.
 XX

PA (ASAH) ASahi KASeT KOGYO KK.
 PA (INOM/) INOMAWARI M.

XX WPI; 1994-269451/33.

XX T cell epitope present in the core protein region of Hepatitis C
 PT virus (HCV) - used for activation of cellular immunity mechanisms

XX Disclosure; Fig 2; 14pp; Japanese.

XX A T-cell stimulating peptide is claimed which is ca. 5-20 amino
 CC acids long and part of the core protein of HCV and is recognised by
 CC and stimulates T-cells. A CD8-positive T-cell stimulating peptide
 CC has the sequence given in AAR56606 (NP-9) or AAR56616 (9MA). A CD4-
 CC positive T-cell stimulating peptide has the sequence given in AAR56609
 CC (NP-12) or AAR56614 (NP-17). Synthetic peptide mixts. (Mix A: NP-1 -
 CC NP-5, Mix B: NP-6 - NP-10, Mix C: NP-11 - NP14, MixD: NP-15 - NP-18)
 CC were used in experiments to evaluate activity.
 CC Peptides 9MA, 9CM and 9CA (AAR56616-18) are variants of peptide NP-9.

XX Sequence 20 AA;

Query Match 100.0%; Score 51; DB 15; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00073;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

Db 1 adlmgypiv 10
 |||||

RESULT 15

AAR84500

ID AAR84500 standard; peptide; 20 AA.

XX AC AAR84500;

XX 06-JAN-1997 (first entry)

XX Hepatitis C virus peptide CORE 21 (residues 121-140).

XX Hepatitis C virus; HCV; immunogen; core region; nucleocapsid;
 KW immunodominant; T cell epitope; vaccine.

XX Hepatitis C virus.

XX WO9512677-A2.

XX 11-MAY-1995.

XX 28-OCT-1994; 94WO-EP03555.

XX 04-NOV-1993; 93EP-0402718.

XX (INNO-) INNOGENETICS NV.

XX Deleys R, Leroux-Roels G, Maertens G;

XX WPI; 1995-193822/25.

XX Hepatitis C Virus immunogenic polypeptide contg. a T-cell
 PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
 PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

XX A series of overlapping peptides (including the present sequence) was
 CC synthesised based on sequences in the core, E1 and E2/NS1 regions of
 CC hepatitis C virus. The peptides were used as antigens in lympho-
 CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

Query Match 100.0%; Score 51; DB 16; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00073;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

Db 11 adlmgypiv 20
 |||||

Search completed: August 23, 2002, 10:00:57
 Job time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:54:30 ; Search time 61.77 Seconds
(without alignments)
3.954 Million cell updates/sec

Title: US-08-854-825-1
Sequence: 51
1 ADLMGYIPLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	100.0	10	1 US-08-214-650-1	Sequence 1, Appl
2	51	100.0	16	2 US-08-986-234-34	Sequence 34, Appl
3	51	100.0	76	1 US-08-324-977-18	Sequence 18, Appl
4	51	100.0	76	2 US-08-384-616-18	Sequence 18, Appl
5	51	100.0	76	2 US-08-904-686A-18	Sequence 18, Appl
6	51	100.0	76	4 US-09-315-850-18	Sequence 18, Appl
7	51	100.0	135	4 US-08-444-818-91	Sequence 91, Appl
8	51	100.0	150	1 US-07-681-703B-16	Sequence 16, Appl
9	51	100.0	150	2 US-08-407-410B-16	Sequence 16, Appl
10	51	100.0	150	2 US-08-485-500-16	Sequence 16, Appl
11	51	100.0	150	5 PCT-US91-02370-16	Sequence 16, Appl
12	51	100.0	154	3 US-08-854-531-2	Sequence 2, Appl
13	51	100.0	154	5 PCT-US95-13552-2	Sequence 2, Appl
14	51	100.0	158	4 US-08-836-075A-66	Sequence 66, Appl
15	51	100.0	169	4 US-08-444-818-93	Sequence 93, Appl
16	51	100.0	190	1 US-07-681-701-16	Sequence 16, Appl
17	51	100.0	190	4 US-08-078-271B-1	Sequence 1, Appl
18	51	100.0	191	2 US-08-290-665A-155	Sequence 155, Appl
19	51	100.0	191	2 US-08-290-665A-156	Sequence 156, Appl
20	51	100.0	191	2 US-08-290-665A-157	Sequence 157, Appl
21	51	100.0	191	2 US-08-290-665A-158	Sequence 158, Appl
22	51	100.0	191	2 US-08-290-665A-159	Sequence 159, Appl
23	51	100.0	191	2 US-08-290-665A-160	Sequence 160, Appl
24	51	100.0	191	2 US-08-290-665A-161	Sequence 161, Appl
25	51	100.0	191	2 US-08-290-665A-162	Sequence 162, Appl
26	51	100.0	191	2 US-08-290-665A-163	Sequence 163, Appl
27	51	100.0	191	2 US-08-290-665A-164	Sequence 164, Appl

28	51	100.0	191	2 US-08-290-665A-165	Sequence 165, App
29	51	100.0	191	2 US-08-290-665A-166	Sequence 166, App
30	51	100.0	191	2 US-08-290-665A-167	Sequence 167, App
31	51	100.0	191	2 US-08-290-665A-168	Sequence 168, App
32	51	100.0	191	2 US-08-290-665A-169	Sequence 169, App
33	51	100.0	191	2 US-08-290-665A-170	Sequence 170, App
34	51	100.0	191	2 US-08-290-665A-171	Sequence 171, App
35	51	100.0	191	2 US-08-290-665A-172	Sequence 172, App
36	51	100.0	191	2 US-08-290-665A-173	Sequence 173, App
37	51	100.0	191	2 US-08-290-665A-174	Sequence 174, App
38	51	100.0	191	2 US-08-290-665A-175	Sequence 175, App
39	51	100.0	191	2 US-08-290-665A-176	Sequence 176, App
40	51	100.0	191	2 US-08-290-665A-187	Sequence 187, App
41	51	100.0	191	2 US-08-290-665A-188	Sequence 188, App
42	51	100.0	191	2 US-08-290-665A-189	Sequence 189, App
43	51	100.0	191	2 US-08-290-665A-190	Sequence 190, App
44	51	100.0	191	2 US-08-290-665A-192	Sequence 192, App
45	51	100.0	191	2 US-08-290-665A-193	Sequence 193, App

ALIGNMENTS

RESULT 1
US-08-214-650-1
; Sequence 1, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; Lymphocyte Responses to Hepatitis C Virus
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-214-650-1

Query Match 100.0%; Score 51; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||||

Db 1 ADLMGYIPLV 10

RESULT 2

US-08-986-234-34
; Sequence 34, Application US/08986234
; Patent No. 5981706
; GENERAL INFORMATION:
; APPLICANT: Wallen, et al.
; TITLE OF INVENTION: Methods for Synthesizing Heat Shock Protein Complexes
; FILE REFERENCE: UNME-0008-1
; CURRENT APPLICATION NUMBER: US/08/986,234
; CURRENT FILING DATE: 1997-12-05
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-08-986-234-34

Query Match 100.0%; Score 51; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00047;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 3 ADLMGYIPLV 12

RESULT 3

US-08-324-977-18
; Sequence 18, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westernman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 23-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; TELECOMMUNICATION INFORMATION: 900703D
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-324-977-18

Query Match 100.0%; Score 51; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 16 ADLMGYIPLV 25

RESULT 4

US-08-384-616-18
; Sequence 18, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westernman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,616
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703B
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-384-616-18

Query Match 100.0%; Score 51; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 16 ADLMGYIPLV 25

RESULT 5

US-08-904-686A-18
Sequence 18, Application US/08904686A
Patent No. 5998130

GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
TITLE OF INVENTION: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, McLeLland &
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,686A
FILING DATE: 01-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/324,977
FILING DATE: 18-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,706
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: McLeLland, Le-Nhung
REGISTRATION NUMBER: 31,541
REFERENCE/DOCKET NUMBER: 900703G
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-904-686A-18

Query Match 100.0%; Score 51; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 16 ADLMGYIPLV 25

RESULT 6

US-09-315-850-18
Sequence 18, Application US/09315850
Patent No. 6217872

GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, McLeLland &
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/315,850
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/904,686
FILING DATE: 01-AUG-1997
APPLICATION NUMBER: US 08/324,977
FILING DATE: 18-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeeland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-315-850-18

Query Match 100.0%; Score 51; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 16 ADLMGYIPLV 25

RESULT 7
US-08-444-818-91
; Sequence 91, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3865
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-91

Query Match 100.0%; Score 51; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 19 ADLMGYIPLV 28

RESULT 8
US-07-681-703B-16
; Sequence 16, Application US/07681703B
; Patent No. 5443965
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/681,703B
; FILING DATE: 05-APR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-681-703B-16

Query Match 100.0%; Score 51; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 131 ADLMGYIPLV 140

RESULT 9
US-08-407-410B-16
; Sequence 16, Application US/08407410B

Patent No. 5843636
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: 350 Cambridge Ave., Suite 100
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,410B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-407-410B-16

Query Match 100.0% Score 51; DB 2; Length 150;
Best Local Similarity 100.0% Pred. No. 0.0057;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 10
US-08-485-500-16
; Sequence 16, Application US/08485500
; Patent No. 5843639
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: 350 Cambridge Ave., Suite 100
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,500
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,410
; FILING DATE:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-500-16

Query Match 100.0% Score 51; DB 2; Length 150;
Best Local Similarity 100.0% Pred. No. 0.0057;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 11
PCT-US91-02370-16
; Sequence 16, Application PC/TUS9102370
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: P.O. BOX 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02370
; FILING DATE: 19910405
; CLASSIFICATION: 435.5
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:

; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US91-02370-16

Query Match 100.0%; Score 51; DB 5; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 12
US-08-854-531-2
; Sequence 2, Application US/08854531
; Patent No. 6025341
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack
; APPLICANT: Tokushige, Katsutoshi
; APPLICANT: Wakita, Takaji
; TITLE OF INVENTION: CHIMERIC HEPATITIS B/HEPATITIS C VIRUS VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6025341 Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,531
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: APOL-0214
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

Query Match 100.0%; Score 51; DB 3; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 13
PCT-US95-13552-2
; Sequence 2, Application PC/TUS9513552
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack
; APPLICANT: Tokushige, Katsutoshi
; APPLICANT: Wakita, Takaji
; APPLICANT: Pachuk, Catherine J.
; APPLICANT: zurawski, Jr., Vincent R.
; APPLICANT: Coney, Leslie R.
; TITLE OF INVENTION: HEPATITIS VIRUS VACCINES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13552
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/318,248
; FILING DATE: 05-OCT-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467,859
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: APOL-0238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-13552-2

Query Match 100.0%; Score 51; DB 5; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 14
US-08-836-075A-66
; Sequence 66, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836.075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-66

Query Match 100.0%; Score 51; DB 4; Length 158;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
Db 131 ADLMGYIPLV 140

RESULT 15
US-08-444-818-93
Sequence 93, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444.818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403.590
FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-93

Query Match 100.0%; Score 51; DB 4; Length 169;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
Db 123 ADLMGYIPLV 132

Search completed: August 23, 2002, 09:57:50
Job time: 200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:54:30 ; Search time 78.82 seconds
(without alignments)
12.191 Million cell updates/sec

Title: US-08-854-825-1
Perfect score: 51
Sequence: 1 ADLMGYIPLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	178	PS0388	genome polyprotein
2	51	100.0	189	S32740	polyprotein - hepa
3	51	100.0	322	JN0265	genome polyprotein
4	51	100.0	369	S21471	genome polyprotein
5	51	100.0	411	PC2060	genome polyprotein
6	51	100.0	411	PC2061	genome polyprotein
7	51	100.0	441	S12707	genome polyprotein
8	51	100.0	492	S41288	genome polyprotein
9	51	100.0	513	PC1284	genome polyprotein
10	51	100.0	513	A44150	structural protein
11	51	100.0	520	JQ1925	polyprotein - hepa
12	51	100.0	523	JQ1926	polyprotein - hepa
13	51	100.0	550	JH0711	genome polyprotein
14	51	100.0	640	JQ1584	genome polyprotein
15	51	100.0	782	S18031	genome polyprotein
16	51	100.0	782	S18032	genome polyprotein
17	51	100.0	782	S19875	genome polyprotein
18	51	100.0	787	PN0677	hypothetical prote
19	51	100.0	876	PC2219	polypeptide - hepa
20	51	100.0	3010	GNWVC	genome polyprotein
21	51	100.0	3010	1 GNWVCJ	genome polyprotein
22	51	100.0	3010	1 A45573	genome polyprotein
23	51	100.0	3010	1 S18030	genome polyprotein
24	51	100.0	3010	1 GNWVTW	genome polyprotein
25	51	100.0	3011	1 GNWVC3	genome polyprotein
26	51	100.0	3011	1 GNWVCH	genome polyprotein
27	51	100.0	3011	1 S40770	genome polyprotein
28	48	94.1	874	2 JQ0883	genome polyprotein
29	48	94.1	874	2 JQ0881	genome polyprotein

ALIGNMENTS

RESULT 1

PS0388

genome polyprotein - hepatitis C virus (isolate GM1) (fragment)

N:Contains: amino end of envelope protein M; carboxyl end of capsid protein C

C:Species: hepatitis C virus

A:Note: host Homo sapiens (man)

C>Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 17-Nov-2000

C:Accession: PS0388

R:Fuchs, K.; Motz, M.; Schreier, E.; Zachoval, R.; Deinhardt, F.; Roggendorf, M.

Gene 103, 163-169, 1991

A:Title: Characterization of nucleotide sequences from European hepatitis C virus iso

A:Reference number: JN0265; MUID:91365241

A:Accession: PS0388

A:Molecule type: genomic RNA

A:Residues: 1-178 <FUC>

A:Cross-references: GB:M61719; NID:G329757; PIDN:AAA45536.1; PID:G387654

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; envelope protein; polyprotein; transmembrane protein

F:1-109/Product: capsid protein C (fragment) #status predicted <CPC>

F:110-178/Product: envelope protein M (fragment) #status predicted <EPM>

F:163-178/Domain: transmembrane #status predicted <TML>

Query Match 100.0%; Score 51; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.022; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 1 ADLMGYIPLV 10

|||||

Db 125 ADLMGYIPLV 134

RESULT 2

S32740

polyprotein - hepatitis C virus (isolate Russian) (fragment)

N:Contains: capsid protein C; envelope protein M

C:Species: hepatitis C virus

C>Date: 19-Mar-1997 #sequence_revision 05-Dec-1998 #text_change 17-Nov-2000

C:Accession: S32740

R:Vassilev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.

submitted to the EMBL Data Library, April 1993

A:Description: Evidence of new HCV variant of European isolate in Russia.

A:Reference number: S32740

A:Accession: S32740

A:Molecule type: genomic RNA

A:Residues: 1-189 <VAS>

A:Cross-references: EMBL:X71407

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; envelope protein; polyprotein

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-189/Product: envelope protein M #status predicted <EPM>

genome polyprotein
genome polyprotein
genome polyprotein
genome polyprotein
rod shape-determin
hypothetical prote
phospho-beta-galac
hypothetical prote
precorrin-3 methyl
hypothetical prote
hypothetical prote
protein FIN21.18
hypothetical prote
probable precorrin
nicotinic acetylch

30 48 94.1 3033 1 JQ1303
31 48 94.1 3033 1 GNWVJ8
32 44 86.3 782 2 S19876
33 41 80.4 3014 1 JC5620
34 38 74.5 340 2 H72325
35 37 72.5 252 2 A70529
36 37 72.5 482 2 JE0395
37 36 70.6 218 2 B86635
38 36 70.6 245 2 T03534
39 36 70.6 297 2 H69609
40 36 70.6 344 2 B84204
41 35 68.6 107 2 AE1083
42 35 68.6 240 2 A96697
43 35 68.6 241 2 T22513
44 35 68.6 242 2 T51115
45 35 68.6 502 1 ACHUA7

```
Query Match          100.0%; Score 51; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
    |||||
Db 131 ADLMGYIPLV 140

RESULT 3
JN0265
genome polyprotein - hepatitis C virus (isolate GM2) (fragments)
N:Contains: amino end of envelope protein M; carboxyl end of capsid protein C; fragment
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 17-Nov-2000
C:Accession: JN0265
R:Fuchs, K.; Motz, M.; Schreier, E.; Zachoval, R.; Deinhardt, F.; Roggendorf, M.
Gene 103, 163-169, 1991
A:Title: Characterization of nucleotide sequences from European hepatitis C virus isolat
A:Reference number: JN0265; MUID:91365241
A:Accession: JN0265
A:Molecule type: genomic RNA
A:Residues: 1-322 <FUC>
A:Cross-references: GB:M61717; GB:M61718
A:Note: the authors translated the codon ACA for residue 198 as Tyr
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein; transmembrane p
F:1-109/Product: capsid protein C (fragment) #status predicted <COR>
F:110-178/Product: envelope protein M (fragment) #status predicted <EPW>
F:163-178/Domain: transmembrane #status predicted <TM1>
F:179-322/Product: major envelope protein E (fragment) #status predicted <ENV>
F:253-269/Domain: transmembrane #status predicted <TM2>
F:191,216/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          100.0%; Score 51; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
    |||||
Db 125 ADLMGYIPLV 134

RESULT 4
S21471
genome polyprotein - hepatitis C virus (fragment)
N:Contains: capsid protein; envelope protein
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
C:Accession: S21471
R:Mogam, W.K.
submitted to the EMBL Data Library, April 1992
A:Reference number: S21471
A:Accession: S21471
A:Molecule type: genomic RNA
A:Residues: 1-369 <MOG>
A:Cross-references: EMBL:X65924; NID:g59466; PIDN:CAA46717.1; PID:g59467
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

Query Match          100.0%; Score 51; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
    |||||
Db 131 ADLMGYIPLV 140
```

```
RESULT 5
PC2060
genome polyprotein N1 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2060
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744
A:Accession: PC2060
A:Molecule type: mRNA
A:Residues: 1-411 <LIJ>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protei
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          100.0%; Score 51; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
    |||||
Db 131 ADLMGYIPLV 140

RESULT 6
PC2061
genome polyprotein N2 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2061
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744
A:Accession: PC2061
A:Molecule type: mRNA
A:Residues: 1-411 <LIJ>
A:Cross-references: GB:L12355; NID:g410169; PIDN:AAA20155.1; PID:g410170
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protei
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          100.0%; Score 51; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
    |||||
Db 131 ADLMGYIPLV 140

RESULT 7
S12707
genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: S12707
R:Takeuchi, K.; Kubo, Y.; Boonmar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G
Nucleic Acids Res. 18, 4626, 1990
A:Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus geno
A:Reference number: S12707; MUID:90356432
A:Accession: S12707
```


A:Molecule type: genomic RNA

A:Residues: 1-441 <TAK>
A:Cross-references: EMBL:D00574; NID:g221656; PIDN:BAA00452.1; PID:g221657
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: polypeptide

Query Match 100.0%; Score 51; DB 2; Length 441;
Best Local Similarity 100.0%; Pred. No. 0.054;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 131 ADLMGYIPLV 140

RESULT 8

S41288

genome polypeptide - hepatitis C virus (fragment)

N:Contains: core protein; envelope protein; NS1 protein

C:Species: hepatitis C virus

C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41288

R:Seelig, R. submitted to the EMBL Data Library, December 1993

A:Reference number: S41288

A:Accession: S41288

A:Molecule type: genomic RNA

A:Residues: 1-492 <SEE>

A:Cross-references: EMBL:X76918

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein; poly

F:1-191/Product: core protein #status predicted <COR>

F:192-372/Product: envelope protein #status predicted <ENV>

F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 100.0%; Score 51; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 131 ADLMGYIPLV 140

RESULT 9

PC1284

genome polypeptide - hepatitis C virus (isolate HC-J4) (fragment)

C:Species: hepatitis C virus

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000

C:Accession: PC1284

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Jpn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116

A:Accession: PC1284

A:Molecule type: genomic RNA

A:Residues: 1-513 <OKA>

A:Cross-references: GB:D00832; NID:g221513; PIDN:BAA00706.1; PID:g221514

C:Superfamily: hepatitis C virus genome polypeptide

Query Match 100.0%; Score 51; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 131 ADLMGYIPLV 140

RESULT 10

A44150

structural protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000

C:Accession: A44150

R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley,

Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992

A:Title: Interaction of immune sera with synthetic peptides corresponding to the stru

A:Reference number: A44150; MUID:92228749

A:Accession: A44150

A:Status: preliminary: not compared with conceptual translation

A:Molecule type: genomic RNA

A:Residues: 1-513 <CHI>

C:Superfamily: hepatitis C virus genome polypeptide

Query Match 100.0%; Score 51; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 131 ADLMGYIPLV 140

RESULT 11

JQ1925

polypeptide - hepatitis C virus (isolate HCV-KF)

N:Contains: C protein; E1 protein; E2/NS1 protein

C:Species: hepatitis C virus

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C:Accession: JQ1925

R:Abe, K.; Inchauspe, G.; Fujisawa, K.

J. Gen. Virol. 73, 2725-2729, 1992

A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated fro

A:Reference number: JQ1925; MUID:93019030

A:Accession: JQ1925

A:Molecule type: mRNA

A:Residues: 1-520 <ABE>

A:Cross-references: DBJ:D10687; NID:g221544; PIDN:BAA01529.1; PID:g221545

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: polypeptide; transmembrane protein

F:1-191/Product: C protein #status predicted <CP>

F:192-383/Product: E1 protein #status predicted <E1>

F:384-520/Product: E2/NS1 protein #status predicted <E2>

Query Match 100.0%; Score 51; DB 2; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 131 ADLMGYIPLV 140

RESULT 12

JQ1926

polypeptide - hepatitis C virus (isolate HCV-476)

N:Contains: C protein; E1 protein; E2/NS1 protein

C:Species: hepatitis C virus

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C:Accession: JQ1926

R:Abe, K.; Inchauspe, G.; Fujisawa, K.

J. Gen. Virol. 73, 2725-2729, 1992

A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated fro

A:Reference number: JQ1925; MUID:93019030

A:Accession: JQ1926

A:Molecule type: mRNA

A:Residues: 1-523 <ABE>

A:Cross-references: DBJ:D10687

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: polypeptide

Query Match 100.0%; Score 51; DB 2; Length 640;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:54:30 ; Search time 37 seconds
(without alignments)
10.465 Million cell updates/sec

Title: US-08-854-825-1
Perfect score: 51
Sequence: 1 ADLMGYIPLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	51	100.0	192	1 POLG_HCV1	P27954 hepatitis c
2	51	100.0	309	1 POLG_HCVH7	P27955 hepatitis c
3	51	100.0	321	1 POLG_HCVH8	P27956 hepatitis c
4	51	100.0	321	1 POLG_HCVTH	P27957 hepatitis c
5	51	100.0	513	1 POLG_HCVJ2	P27959 hepatitis c
6	51	100.0	520	1 POLG_HCVH4	Q01404 hepatitis c
7	51	100.0	520	1 POLG_HCVHK	Q01403 hepatitis c
8	51	100.0	3010	1 POLG_HCVBK	P26663 h genome po
9	51	100.0	3010	1 POLG_HCVJA	P26662 h genome po
10	51	100.0	3010	1 POLG_HCVJT	Q00269 h genome po
11	51	100.0	3010	1 POLG_HCVTW	P29846 h genome po
12	51	100.0	3011	1 POLG_HCVI	P26664 h genome po
13	51	100.0	3011	1 POLG_HCVH	P27958 h genome po
14	48	94.1	737	1 POLG_HCVJ5	P27960 hepatitis c
15	48	94.1	737	1 POLG_HCVJ7	P27961 hepatitis c
16	48	94.1	3033	1 POLG_HCVJ6	P26660 h genome po
17	48	94.1	3033	1 POLG_HCVJ8	P26661 h genome po
18	36	70.6	218	1 YAFI_LACLA	O5cjb5 lactococcus
19	36	70.6	297	1 CTAG_BACSU	Q34329 bacillus su
20	35	68.6	240	1 Y136_ARATH	Q9fyf7 arabidopsis
21	35	68.6	241	1 YX70_CAEEL	O17883 caenorhabdi
22	35	68.6	499	1 ACH7_BOVIN	P54131 bos taurus
23	35	68.6	502	1 ACH7_HUMAN	P36544 homo sapien
24	35	68.6	604	1 NU5M_ORNAN	Q36459 ornithorhyn
25	35	68.6	768	1 PARC_NEIGO	P48374 neisseria g
26	34	66.7	196	1 CLPP_HELPY	P46156 helicobacte
27	34	66.7	255	1 CAPC_STAAU	P39852 staphylococ
28	34	66.7	479	1 BGLA_ECOLI	Q46829 escherichia
29	34	66.7	606	1 NU5M_FELCA	P48921 felis silve
30	33	64.7	177	1 MLF2_MALFU	P56577 malassezia
31	33	64.7	248	1 YFGE_ECOLI	P76570 escherichia
32	33	64.7	281	1 PH4H_CHRVO	P30967 chromobacte
33	33	64.7	372	1 3BH1_RAT	P22071 r 3 beta-hy

34 33 64.7 452 1 PH4H_DROME P17276 drosophila
35 33 64.7 501 1 C72N_ARATH Q91tm0 arabidopsis
36 33 64.7 1239 1 CHS5_USTMA O13394 ustilago ma
37 32 62.7 195 1 CLPP_HELPJ Q92150 helicobacte
38 32 62.7 210 1 HUPD_RHOCA Q03004 rhodobacter
39 32 62.7 226 1 YN74_YEAST Q05024 saccharomyc
40 32 62.7 311 1 CYSM_BACSU Q34476 bacillus su
41 32 62.7 351 1 RPOA_TREPA O83242 treponema p
42 32 62.7 385 1 GBAL_SOYBN P49084 glycine max
43 32 62.7 433 1 COBB_PSEDE P21632 pseudomonas
44 32 62.7 435 1 CSBX_BACSU O05390 bacillus su
45 32 62.7 446 1 CITN_KLEPN P31602 klebsiella

ALIGNMENTS

RESULT 1
ID POLG_HCV1 STANDARD: PRT: 192 AA.
AC P27954;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contents: Matrix protein (Envelope protein M);
DE Major envelope protein E] (Fragment).
OS Hepatitis C virus (isolate EC1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11107;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; Pubmed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:843-848(1991).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53135; CAA37295.1; -;
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC Pfam; PF01542; HCV_core: 1.
CC Pfam; PF01539; HCV_env: 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane.
CC NON_TER 1
CC CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 >192 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 20315 MW; 98E488F4C335A84C CRC64;

Query Match 100.0%; Score 51; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.013;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
|||||
Db 15 ADLMGYIPLV 24

RESULT 2

POLG_HCVH7 STANDARD; PRT; 309 AA.
AC P27955;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCV27) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RL Virology 180:842-848(1991).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53133; CAA37293.1; -
CC InterPro; IPR002531; HCV_NSI.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC ProDom; PD186062; HCV_NSI; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 63
FT CHAIN 54 255
FT CHAIN 256 >309
FT CARBOHYD 68 68
FT CARBOHYD 81 81
FT CARBOHYD 106 106
FT CARBOHYD 177 177
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 302 302
FT NON_TER 309
SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ADLMGYIPLV 10
|||||
Db 3 ADLMGYIPLV 12

RESULT 3

POLG_HCVH8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RL Virology 180:842-848(1991).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53131; CAA37291.1; -
CC InterPro; IPR002531; HCV_NSI.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC ProDom; PD186062; HCV_NSI; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 76
FT CHAIN 77 267
FT CHAIN 268 >321
FT CARBOHYD 80 80
FT CARBOHYD 93 93
FT CARBOHYD 118 118
FT CARBOHYD 189 189
FT CARBOHYD 301 301
FT CARBOHYD 307 307
FT CARBOHYD 314 314
FT NON_TER 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ADLMGYIPLV 10
|||||
Db 15 ADLMGYIPLV 24

RESULT 4

POLG_HCVTH STANDARD; PRT; 321 AA.
ID POLG_HCVTH

```
AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DE 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53134; CAA37294.1; -
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002521; HCV_Core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW NON_TER 1
FT CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
DB 15 ADLMGYIPLV 24

RESULT 5
ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10074; BAA00968.1; -
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002519; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW INIT_MET 1
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
DB 131 ADLMGYIPLV 140
```

```
RESULT 6
POLG_HCVH4 STANDARD; PRT; 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1) (Fragment)].
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10688; BAA01530.1; -.
DR InterPro; IPR002531; HCV_Ns1.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_Ns1; 1.
DR ProDom; PD186062; HCV_Ns1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADLMGYIPLV 10
```

```
|||||||
Db 131 ADLMGYIPLV 140

RESULT 7
POLG_HCVHK STANDARD; PRT; 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1) (Fragment)].
OS Hepatitis C virus (isolate HCV-KF) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10687; BAA01529.1; -.
DR PIR; J01925; J01925.
DR InterPro; IPR002531; HCV_Ns1.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_Ns1; 1.
DR ProDom; PD186062; HCV_Ns1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 100.0%; Score 51; DB 1; Length 520;
```


Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
IDB 131 ADLMGYIPLV 140

RESULT 8

POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RA "Structure and organization of the hepatitis C virus genome isolated
RA from human carriers.";
PL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RA "Non-structural protein 3 of hepatitis C virus inhibits
RA phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomsky Z.;
RA "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RA trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RA "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RA virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; M58335; AAA72945.1; -;
DR PIR; A38465; GNWVTC.
DR PDB; 1A1Q; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR PDB; 1NS3; 08-APR-98.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).

```

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 100.0%; Score 51; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
   |||||
Db 131 ADLMGYIPLV 140

RESULT 9
POLG_HCVJA STANDARD; PRT; 3010 AA.
ID POLG_HCVJA
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/Helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=11116;
RX [1]
RP SEQUENCE FROM N.A.
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
[2]
RN DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC 1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC 1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC 1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND MRNA.
CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D90208; BAAL4233.1; -.
CC PIR; A39253; GNWVCJ.
CC HSP; P26663; 1JXP.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.

```

```

DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3010 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 536 536
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327017 MW; AA993794F6DB185 CRC64;

```

Query Match 100.0%; Score 51; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

Db 131 ADLMGYIPLV 140

```
RESULT 10
POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D11168; BAA01943.1; -.
CC PIR; A45573; A45573.
CC HSP; P26663; LJXP.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002866; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRP.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
```

```
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
```

Query Match 100.0%; Score 51; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10

|||||

Db 131 ADLMGYIPLV 140

RESULT 11

POLG_HCVTW STANDARD; PRT; 3010 AA.

ID POLG_HCVTW

AC P29846;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
RN NCBI_TaxID=131645;
RP [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=92230206; PubMed=1314449;
RT Chen P.J., Lin M.H., Tal K.F., Liu P.C., Lin C.J., Chen D.S.;
"The Taiwanese hepatitis C virus genome: sequence determination and
mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the p6
position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR PIR; A40244; GNWVTV.
DR HSP; P26663; LUXP.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_Ns1.
DR InterPro; IPR002518; HCV_Ns2.
DR InterPro; IPR004109; HCV_Ns3.
DR InterPro; IPR000745; HCV_Ns4a.
DR InterPro; IPR001490; HCV_Ns4b.
DR InterPro; IPR002868; HCV_Ns5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_Ns1; 1.
DR Pfam; PF01538; HCV_Ns2; 1.
DR Pfam; PF02907; HCV_Ns3; 1.
DR Pfam; PF01006; HCV_Ns4a; 1.
DR Pfam; PF01001; HCV_Ns4b; 1.
DR Pfam; PF01506; HCV_Ns5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_Ns1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369
FT ACT_SITE 1083
FT ACT_SITE 1107
FT ACT_SITE 1165
FT NP_BIND 1230
FT SITE 1316
FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 233
FT CARBOHYD 234
FT CARBOHYD 250
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 448
FT CARBOHYD 532
FT CARBOHYD 540
FT CARBOHYD 556
FT CARBOHYD 576
FT CARBOHYD 623
FT CARBOHYD 645
FT CARBOHYD 2041
FT CARBOHYD 2077
FT CARBOHYD 2240
FT CARBOHYD 2529
FT CARBOHYD 2788
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 100.0%; Score 51; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
DB 131 ADLMGYIPLV 140

RESULT 12
POLG_HCV1 STANDARD; PRT: 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
"Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the p6
position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M62321; AAA5676.1; -;
 DR PIR; A39166; GNMVC3.
 DR HSP; P27958; IHEI.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR00745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RdRP; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00492; HELIC3; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 100.0%; Score 51; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ADLMGYIPLV 10
 Db 131 ADLMGYIPLV 140
 RESULT 13
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.46)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.


```
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 94.1%; Score 48; DB 1; Length 737;
Best Local Similarity 90.0%; Pred. No. 0.17;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
DB 131 ADLMGYIPV 140
|||||||:|

RESULT 15
POLG_HCVJ7
ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment);
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed-1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 186:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR ENBL; D10077; BAA00971.1; -
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR Transmembrane; Glycoprotein; Coat protein; Envelope protein;
KW Polyprotein; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 94.1%; Score 48; DB 1; Length 737;
Best Local Similarity 90.0%; Pred. No. 0.17;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
DB 131 ADLMGYIPV 140
|||||||:|
```

Search completed: August 23, 2002, 09:55:17
Job time: 47 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:00 ; Search time 139.83 Seconds
(without alignments)
12.372 Million cell updates/sec

Title: US-08-854-825-1
Perfect score: 51
Sequence: 1 ADLMGYIPLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mic:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51	100.0	150	12 Q68843	Q68843 hepatitis c
2	51	100.0	150	12 Q68867	Q68867 hepatitis c
3	51	100.0	150	12 Q68841	Q68841 hepatitis c
4	51	100.0	151	12 Q68961	Q68961 hepatitis c
5	51	100.0	163	12 P89094	P89094 hepatitis c
6	51	100.0	170	12 Q81276	Q81276 hepatitis c
7	51	100.0	173	12 Q9E8Y1	Q9E8Y1 hepatitis c
8	51	100.0	178	12 Q81275	Q81275 hepatitis c
9	51	100.0	178	12 Q03729	Q03729 hepatitis c
10	51	100.0	186	12 Q09738	Q09738 hepatitis c
11	51	100.0	186	12 Q09739	Q09739 hepatitis c
12	51	100.0	189	12 Q81295	Q81295 hepatitis c
13	51	100.0	190	12 Q68873	Q68873 hepatitis c
14	51	100.0	190	12 Q81244	Q81244 hepatitis c
15	51	100.0	191	12 Q9PXE6	Q9PXE6 hepatitis c
16	51	100.0	191	12 P90247	P90247 hepatitis c

ALIGNMENTS

RESULT 1

Q68843 ID Q68843 PRELIMINARY; PRT; 150 AA.
AC Q68843;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
GN HCV CORE.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India : Identification of a new subtype.";
RL J. Med. Virol. 48:191-198(1996).
DR EMBL; X91297; CAA62671.1; -;
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
FT NON_TER 1
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 16678 MW; 251AC56249B26432 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
Db 128 ADLMGYIPLV 137

```
RESULT 2
Q68867 ID Q68867 PRELIMINARY; PRT; 150 AA.
AC Q68867;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
GN HCV CORE.
OS Hepatitis C virus type 3.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=40363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IND 674;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India :
RT Identification of a new subtype.";
RL J. Med. Virol. 48:191-198(1996).
DR EMBL; X91300; CAA62674.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16509 MW; DE75E728FF39B2B3 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 128 ADLMGYIPLV 137

RESULT 3
Q68841 ID Q68841 PRELIMINARY; PRT; 150 AA.
AC Q68841;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
GN HCV CORE.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IND 1358;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India :
RT Identification of a new subtype.";
RL J. Med. Virol. 48:191-198(1996).
DR EMBL; X91304; CAA62678.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16685 MW; 71E3A2B5C196994E CRC64;

Query Match 100.0%; Score 51; DB 12; Length 150;
```

```
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 128 ADLMGYIPLV 137

RESULT 4
Q68961 ID Q68961 PRELIMINARY; PRT; 151 AA.
AC Q68961;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
GN C2.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95270681; PubMed=7751366;
RA Feucht H.H., Zoellner B., Polywka S., Laufs R.;
RT "Study on reliability of commercially available hepatitis C virus
RT antibody tests.";
RL J. Clin. Microbiol. 33:620-624(1995).
DR EMBL; X78951; CAA55548.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 1
FT NON_TER 151
SQ SEQUENCE 151 AA; 16584 MW; 0DBB1148B3581D6D CRC64;

Query Match 100.0%; Score 51; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
Db 120 ADLMGYIPLV 129

RESULT 5
P89094 ID P89094 PRELIMINARY; PRT; 163 AA.
AC P89094;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96422679; PubMed=8825296;
RA Hitomi Y., McDonnell W.M., Baker J.R., Jr., Askari F.K.;
RT "High efficiency prokaryotic expression and purification of a portion
RT of the hepatitis C core protein and analysis of the immune response to
RT recombinant protein in BALB/c mice.";
RL Viral Immunol. 8:109-119(1995).
DR EMBL; S83169; AAB46894.2; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 163
```

SQ SEQUENCE 163 AA; 17917 MW; D74420694E11A8E4 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 163;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 6

ID Q81276 PRELIMINARY; PRT; 170 AA.
AC Q81276;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95186115; PubMed=7765873;
RA Seki M., Honda Y., Kondo J., Fukuda K., Ohta K., Sugimoto J.,
RA Yamada E.;
RT "Effective production of the hepatitis C virus core antigen having
high purity in *Escherichia coli*.";
RL J. Biotechnol. 38:229-241(1995).
DR EMBL; D30614; BAA06304.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 170 170
SQ SEQUENCE 170 AA; 18622 MW; 7EA0168530E6BBEF CRC64;

Query Match 100.0%; Score 51; DB 12; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 7

ID Q9E8Y1 PRELIMINARY; PRT; 173 AA.
AC Q9E8Y1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIV-1;
RA Jha J., Arankalle V.A.;
RT "Phylogenetic Analysis of Indian HCV Isolates.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF134734; AAG091115.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.

FT NON_TER 173 173
SQ SEQUENCE 173 AA; 18979 MW; C1F650CDFE4968C2 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 173;
Best Local Similarity 100.0%; Pred. No. 0.028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 8

ID Q81275 PRELIMINARY; PRT; 178 AA.
AC Q81275;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91365241; PubMed=1653756;
RA Fuchs K., Motz M., Schreier E., Zachoval R., Deinhardt F.,
RA Roggendorf M.;
RT "Characterization of nucleotide sequences from european hepatitis C
virus isolates.";
RL Gene 103:163-169(1991).
DR EMBL; M61719; AAA45536.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
FT NON_TER 1 1
FT NON_TER 178 178
SQ SEQUENCE 178 AA; 19463 MW; 4ABDCD20895FC87F CRC64;

Query Match 100.0%; Score 51; DB 12; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 125 ADLMGYIPLV 134

RESULT 9

ID Q03729 PRELIMINARY; PRT; 178 AA.
AC Q03729;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91365241; PubMed=1653756;
RA Fuchs K., Motz M., Schreier E., Zachoval R., Deinhardt F.,
RA Roggendorf M.;
RT "Characterization of nucleotide sequences from European hepatitis C
virus isolates.";
RL Gene 103:163-169(1991).
DR EMBL; M61718; AAA45535.1; -.

DR InterPro: IPR002522; HCV_capsid.
DR EMBL: U94723; AAB51538.1; -.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Core protein; Polyprotein.
FT NON_TER 1
SQ SEQUENCE 178 AA; 19483 MW; 93718C60895ADC69 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 125 ADLMGYIPLV 134

RESULT 10
O09738 PRELIMINARY; PRT; 186 AA.
ID O09738;
AC O09738;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NUCLEOCAPSID PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Lerat H., Rumin S., Habersetzer F., Berby F., Traub M.-A., Trepo C.,
RA Inchauspe G.;
RT "Genotype influences the in-vitro tropism of Hepatitis C virus genomic
sequences in hematopoietic cells.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U94722; AAB51537.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
FT NON_TER 186
SQ SEQUENCE 186 AA; 20455 MW; 85EABEB33E71B372 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 11
O09739 PRELIMINARY; PRT; 186 AA.
ID O09739;
AC O09739;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NUCLEOCAPSID PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Lerat H., Rumin S., Habersetzer F., Berby F., Traub M.-A., Trepo C.,
RA Inchauspe G.;
RT "Genotype influences the in-vitro tropism of Hepatitis C virus genomic
sequences in hematopoietic cells.";

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U94723; AAB51538.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
FT NON_TER 186
SQ SEQUENCE 186 AA; 20214 MW; E73DF6878E18CBB5 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 131 ADLMGYIPLV 140

RESULT 12
Q81295 PRELIMINARY; PRT; 189 AA.
ID Q81295;
AC Q81295;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CORE PROTEIN/E1 PROTEIN (FRAGMENT).
OS Hepatitis C virus type 4.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=33745;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAR4/1205;
RA Stuyver L., Fretz C., Jeannel D.;
RT "Hepatitis C virus infection in a rural population in Central African
Republic.";
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: L36439; AAA45537.1; -.
FT NON_TER 189
SQ SEQUENCE 189 AA; 19920 MW; 7CEC828F6A09C398 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 189;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | | | | | | |
Db 5 ADLMGYIPLV 14

RESULT 13
Q68873 PRELIMINARY; PRT; 190 AA.
ID Q68873;
AC Q68873;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94180929; PubMed=8133850;
RA Vassiliev V.B., Viazov S.O., Kotova E.Y., Nosikov V.V.;
RT "Determination of the nucleotide sequence of the Russian variant of
the hepatitis C virus.";
RL Mol. Gen. Mikrobiol. Virusol. 1:33-37(1994).
DR EMBL: X71407; CAA50531.1; -.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
FT NON_TER 1 190
SQ SEQUENCE 190 AA; 20737 MW; E34D2A3AAE167C87 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | |
Db 131 ADLMGYIPLV 140

RESULT 14
Q81244
ID Q81244 PRELIMINARY; PRT; 190 AA.
AC Q81244;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE (BE95) CORE PROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BE95;
RX MEDLINE=95023999; PubMed=7524083;
RA Stuyver L., van Arnhem W., Wyseur A., Hernandez F., Delaporte E.,
Ra Maertens G.;
RT "Classification of hepatitis C viruses based on phylogenetic analysis
of the envelope 1 and nonstructural 5B regions and identification of
five additional subtypes.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:10134-10138(1994).
DR EMBL: L29582; AAA65791.1; -.
FT NON_TER 1 190
FT NON_TER 190 190
SQ SEQUENCE 190 AA; 19776 MW; 2BE966837CAB33D5 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | |
Db 4 ADLMGYIPLV 13

RESULT 15
Q9PXE6
ID Q9PXE6 PRELIMINARY; PRT; 191 AA.
AC Q9PXE6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE CORE PROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96029410; PubMed=7557851;
RA Iwata K., Wakita T., Okumura A., Yoshioka K., Takayanagi M.,
RA Wands J.R., Kakumu S.;
RT "Interferon gamma production by peripheral blood lymphocytes to

RT hepatitis C virus core protein in chronic hepatitis C infection.";
RL Hepatology 22:1057-1064(1995).
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
SQ SEQUENCE 191 AA; 20835 MW; 93CA960134EB8A51 CRC64;

Query Match 100.0%; Score 51; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLMGYIPLV 10
| | | | |
Db 131 ADLMGYIPLV 140

Search completed: August 23, 2002, 10:21:58
Job time: 1498 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:00:57 ; Search time 179.72 Seconds
(without alignments)
6.180 Million cell updates/sec

Title: US-08-854-825-2

Perfect score: 46

Sequence: 1 LIALLSCLTV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	100.0	10	15	AA61498
2	46	100.0	10	15	AA61498
3	46	100.0	10	16	AA73105
4	46	100.0	10	16	AA78942
5	46	100.0	10	16	AA84571
6	46	100.0	10	18	AA39559
7	46	100.0	10	20	AA710482
8	46	100.0	10	20	AA710220
9	46	100.0	15	22	AAJ03042
10	46	100.0	15	22	AAJ03056
11	46	100.0	15	22	AAJ03088
					AAJ03348
					Peptide fragment (
					Antigen fragment 2
					HCV core 178-187 c
					Cytotoxic T-cell e
					HPV16 E7 peptide (
					HLA class I motif
					T cell epitope/MHC
					Hepatitis C virus
					Hepatitis C virus
					Hepatitis C virus
					Hepatitis C virus

12	46	100.0	15	22	AAJ03362	Hepatitis C virus
13	46	100.0	15	22	AAJ03394	Hepatitis C virus
14	46	100.0	20	16	AA84501	Hepatitis C virus
15	46	100.0	135	21	AA18534	Protein encoded by
16	46	100.0	145	13	AA41082	Non-A non-B heptat
17	46	100.0	149	17	AA96527	Hepatitis C virus
18	46	100.0	149	17	AA96537	Hepatitis C virus
19	46	100.0	149	17	AA96540	Hepatitis C virus
20	46	100.0	149	17	AA96541	Hepatitis C virus
21	46	100.0	149	17	AA96542	Hepatitis C virus
22	46	100.0	149	17	AA96544	Hepatitis C virus
23	46	100.0	166	13	AA30062	HCV E1 (envelope p
24	46	100.0	166	20	AA22020	HCV E1 peptide seq
25	46	100.0	166	20	AAW75481	Hepatitis C virus
26	46	100.0	191	14	AA44010	Hepatitis C virus
27	46	100.0	191	15	AA63332	Hepatitis C virus
28	46	100.0	191	15	AA63331	Hepatitis C virus
29	46	100.0	191	15	AA63333	Hepatitis C virus
30	46	100.0	191	17	AA92936	Hepatitis C virus
31	46	100.0	191	17	AA92937	Hepatitis C virus
32	46	100.0	191	17	AA92938	Hepatitis C virus
33	46	100.0	191	17	AA92939	Hepatitis C virus
34	46	100.0	191	17	AA92940	Hepatitis C virus
35	46	100.0	191	17	AA92941	Hepatitis C virus
36	46	100.0	191	17	AA92951	Hepatitis C virus
37	46	100.0	191	17	AA92972	Hepatitis C virus
38	46	100.0	191	17	AA92973	Hepatitis C virus
39	46	100.0	191	17	AA92976	Hepatitis C virus
40	46	100.0	191	17	AA92977	Hepatitis C virus
41	46	100.0	191	17	AA92978	Hepatitis C virus
42	46	100.0	191	19	AAW37125	Hepatitis C virus
43	46	100.0	191	21	AA94411	Human hepatitis C
44	46	100.0	191	21	AA92997	Hepatitis C virus
45	46	100.0	191	21	AA92999	Hepatitis C virus

ALIGNMENTS

RESULT 1
ID AA61498 standard; peptide; 10 AA.
AC AA61498;
XX
DT 11-MAY-1995 (first entry)
XX
DE Peptide fragment (1.0884) of HCV binds HLA-A2.1.
XX
KW antigen; epitope; immunogenic target protein; PSA; HBVC; HBVs; EBV;
KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
KW human immunodeficiency virus; human papilloma virus; p53; c-ERB2;
KW MAGE-1; melanoma antigen-1; core antigen; surface antigen;
KW pharmaceutical composition; in vivo; ex vivo; therapeutic;
KW diagnostic; MHC class I molecule; major histocompatibility complex;
KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen.
XX
OS Hepatitis C virus.
XX
PN WO9420127-A.
XX
PD 15-SEP-1994.
XX
PF 04-MAR-1994; 94WO-US02353.
XX
PR 05-MAR-1993; 93US-0027146.
PR 04-JUN-1993; 93US-0073205.
PR 29-NOV-1993; 93US-0159184.
XX
PA (CYTE-) CYTEL CORP.
XX
PI Grey HM, Kast WM, Sette A, Sidney J;
XX

DR WPI; 1994-302678/37.
 XX Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
 PT for treatment or prophylaxis of cancer, virus infection or
 PT autoimmune diseases.
 XX
 XX Example 5; Page 107; 138pp; English.
 XX
 CC AAR59496-R61666 are immunogenic 10mer peptides that contain a HLA-A2.1
 CC binding motif. These peptides bind HLA-A2.1 and have a binding
 CC affinity of at least 1% as compared to a reference peptide (AAR71293).
 CC AAR61498 has an IC50 of 0.61 and the sequence occurs at position 178
 CC in the HCV CORE protein. The peptides of the invention can induce
 CC cytotoxic T lymphocytes which can react with target cells. They can
 CC be used for the treatment or prophylaxis of cancer, eg. prostate
 CC cancer or lymphoma, etc.
 XX
 XX Sequence 10 AA;
 SQ
 Query Match 100.0%; Score 46; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLALLSCLTV 10
 Db | | | | | | | | | |
 1 llallscstv 10
 RESULT 2
 AAR73105
 ID AAR73105 standard; peptide; 10 AA.
 XX AC
 XX AAR73105;
 XX
 DT 16-JUN-1995 (first entry)
 XX
 DE Antigen fragment 2 from HCV has binding affinity for HLA-2.1.
 XX
 KW antigen; epitope; immunogenic target protein; PSA; HBVc; EBV;
 KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
 KW human immunodeficiency virus; human papilloma virus; p53; C-ERB2;
 KW MAGE-1; melanoma antigen-1; core antigen; surface antigen;
 KW pharmaceutical composition; in vivo; ex vivo; therapeutic;
 KW diagnostic; MHC class I molecule; major histocompatibility complex;
 KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;
 KW algorithm prediction; MBP; CMV; cytomegalovirus; HSV;
 KW herpes simplex virus; influenza A; M1.
 XX
 OS Hepatitis C virus.
 XX
 PN W09420127-A.
 XX
 PD 15-SEP-1994.
 XX
 PF 04-MAR-1994; 94WO-US02353.
 XX
 PR 05-MAR-1993; 93US-0027146.
 XX
 PR 04-JUN-1993; 93US-0073205.
 XX
 PR 29-NOV-1993; 93US-0159184.
 XX
 PA (CYTE-) CYTEL CORP.
 XX
 PI Grey HM, Kast WM, Settle A, Sidney J;
 XX
 XX WPI; 1994-302678/37.
 DR
 XX Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
 PT for treatment or prophylaxis of cancer, virus infection or
 PT autoimmune diseases.
 XX
 XX Disclosure; Page 90; 138pp; English.

CC AAR73058-121 are potential peptide binders of HLA-A2.1 motif. Using
 CC motifs disclosed in the invention, these peptides were screened for
 CC further motifs. Only peptides with binding affinity of at least 1%
 CC (binding affinity is expressed as an IC50 value) as compared to the
 CC standard peptide (AAR71293) in assays. This peptide from hepatitis
 CC C virus has an binding value of 0.6100. The peptides of the invention
 CC can induce cytotoxic T lymphocytes which can react with target cells.
 CC They can be used for the treatment or prophylaxis of cancer, eg.
 CC prostate cancer or lymphoma, etc.
 XX
 XX Sequence 10 AA;
 SQ
 Query Match 100.0%; Score 46; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLALLSCLTV 10
 Db | | | | | | | | | |
 1 llallscstv 10
 RESULT 3
 AAR78942
 ID AAR78942 standard; peptide; 10 AA.
 XX AC
 XX AAR78942;
 XX
 DT 01-APR-1996 (first entry)
 XX
 DE HCV core 178-187 cytotoxic T lymphocyte epitope.
 XX
 KW HCV core 178-187; cytotoxic T; CTL; epitope; helper T; HTL; cell;
 KW lymphocyte; antigens; treatment; disease prevention; hepatitis C;
 KW non-A; non-B.
 XX
 OS Hepatitis C virus.
 XX
 PN W09522317-A1.
 XX
 PD 24-AUG-1995.
 XX
 PF 16-FEB-1995; 95WO-US02121.
 XX
 PR 16-FEB-1994; 94US-0197484.
 XX
 PA (CYTE-) CYTEL CORP.
 XX
 PI Ceut RW, Grey H, Settle AD, Vitiello MA;
 XX
 XX WPI; 1995-302545/39.
 DR
 XX Compsn. inducing cytotoxic T lymphocyte response to pref. viral,
 PT bacterial, parasitic or tumour antigens - useful in the treatment
 PT and prevention of diseases associated with the antigen e.g.
 PT hepatitis B
 XX
 XX Example 12; Page 70; 109pp; English.
 PS
 CC A compsn. which induces a cytotoxic T lymphocyte (CTL) response to
 CC a hepatitis C virus (HCV) antigen (Ag) in a mammal comprises, a
 CC HCV CTL Ag response inducing peptide (i.e. AAR78941-R78955) and a
 CC lipid conjugated helper T cell inducing peptide. The compsn. is
 CC useful in the treatment and prevention of hepatitis C.
 XX
 XX Sequence 10 AA;
 SQ
 Query Match 100.0%; Score 46; DB 16; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLALLSCLTV 10


```

Db      1 llllscitv 10
RESULT  4
AAR84571
ID    AAR84571 standard; peptide; 10 AA.
XX    AC
XX    AC
XX    DT 25-APR-1996 (first entry)
XX    DE Cytotoxic T-cell epitope, aa 178-187 of HCV-1 core region.
XX    KW Hepatitis C virus; HCV; epitope; vaccine; immunogen.
XX    OS Hepatitis C virus.
XX    PN W09525122-A1.
XX    XX
XX    PD 21-SEP-1995.
XX    PF 16-MAR-1995; 95WO-US03224.
XX    PR 17-MAR-1994; 94US-0214650.
XX    PA (SCRI ) SCRIPPS RES INST.
XX    PI Cerny A, Chisari FV;
XX    DR WPI; 1995-336941/43.
XX    PT Novel molecule comprising a cytotoxic T cell epitope - used to
XX    PT vaccinate against hepatitis C viral infection
XX    PS Claim 1; Page 60; 85pp; English.
XX    CC AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides
XX    CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
XX    CC were tested for peptide specific cytotoxic T-cell activity. The
XX    CC peptides AAR84570-77 were found to have substantial homology with a T-
XX    CC cell epitope and are useful in vaccines against HCV infection.
XX    SQ Sequence 10 AA;

Query Match 100.0%; Score 46; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy      1 LLALLSCLTV 10
Db      1 llllscitv 10

RESULT  5
AAW39559
ID    AAW39559 standard; peptide; 10 AA.
XX    AC
XX    AC AAW39559;
XX    DT 11-JUN-1998 (first entry)
XX    DE HPV16 E7 peptide (pos. 82-90) capable of binding to HLA-A*0201.
XX    KW T cell epitope; immune response; human leukocyte antigen; HLA Class I;
XX    KW vaccine; immunogenic; major histocompatibility complex; MHC; B cell;
XX    KW disease; anti-tumour; anti-viral.
XX    OS Human papillomavirus type 16.
XX    PN W09741440-A1.
XX    XX

Pd      06-NOV-1997.
XX    XX
XX    PF 28-APR-1997; 97WO-NL00229.
XX    PR 23-DEC-1996; 96EP-0203670.
XX    PR 26-APR-1996; 96EP-0201145.
XX    PA (OYLE-) RIJKSUNIV LEIDEN.
XX    PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX    PI Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;
XX    DR WPI; 1997-549891/50.
XX    PT Method of selecting T cell peptide epitope(s) - by measuring the
XX    PT stability of HLA class I-peptide complexes on intact B cells
XX    PS Example 2; Page 63; 109pp; English.
XX    CC Peptides AAW39430-W39734 are used in a novel method for the selection of
XX    CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
XX    CC method involves the identification of peptide sequences capable of
XX    CC binding to an HLA (human leukocyte antigen) class I molecule and
XX    CC measuring the binding of this epitope peptide to the HLA class I
XX    CC peptide. The stability of binding of the peptide and MHC (major
XX    CC histocompatibility complex) class I molecule is measured on intact human
XX    CC B cells carrying the MHC molecule at their cell surfaces. The method can
XX    CC be used to select peptide epitopes for generating vaccines against a
XX    CC disease associated with the polypeptide, e.g. cancers or AIDS. The
XX    CC peptide epitopes are especially T-cell peptide epitopes with strong
XX    CC anti-tumour and anti-viral immune responses. Peptides AAW39557-W39559 are
XX    CC conserved human papilloma virus type 16 (HPV16) E6 derived peptides used
XX    CC in an assay to determine immunogenicity and binding affinity to
XX    CC HLA-A*0201.
XX    SQ Sequence 10 AA;

Query Match 100.0%; Score 46; DB 18; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy      1 LLALLSCLTV 10
Db      1 llllscitv 10

RESULT  6
AAV10482
ID    AAV10482 standard; Peptide; 10 AA.
XX    AC
XX    AC AAV10482;
XX    DT 12-MAY-1999 (first entry)
XX    DE HLA Class I motif peptide SEQ ID NO:412.
XX    KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
XX    KW immunisation; tumour; infectious disease; immunotherapy; cancer;
XX    KW malignant melanoma; viral disease; hepatitis; AIDS.
XX    OS Synthetic.
XX    OS Hepatitis C virus.
XX    PN W09902183-A2.
XX    PD 21-JAN-1999.
XX    XX
XX    PF 10-JUL-1998; 98WO-US14289.
XX    PR 10-DEC-1997; 97US-0988320.
XX    PR 10-JUL-1997; 97CA-2209815.
XX    XX

```


CC present sequence is an epitope used in the disclosure of the invention.

```
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 46; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 5 llallscitv 14

RESULT 9
AAJ03056
ID AAJ03056 standard; Peptide; 15 AA.
XX
AC AAJ03056;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3047.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Cellis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
PS Disclosure; Page 174; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 46; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 4 llallscitv 13

RESULT 11
AAJ03348
ID AAJ03348 standard; Peptide; 15 AA.
XX
AC AAJ03348;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3339.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Cellis E, Kubo RT, Grey HM;
XX
```

DR WPI; 2001-308046/32.
XX A new composition useful as a vaccines against hepatitis C virus -
XX
XX Disclosure; Page 177; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 15 AA;
SQ

Query Match 100.0%; Score 46; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
DB 5 LLALLSCLTV 14

RESULT 12
AAJ03362
ID AAJ03362 standard; Peptide; 15 AA.
XX
AC AAJ03362;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3353.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIW-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Disclosure; Page 177; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 15 AA;
SQ

Query Match 100.0%; Score 46; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
DB 5 LLALLSCLTV 14

RESULT 13
AAJ03394
ID AAJ03394 standard; Peptide; 15 AA.
XX
AC AAJ03394;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3385.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIW-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Disclosure; Page 177; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 15 AA;
SQ

Query Match 100.0%; Score 46; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
DB 4 LLALLSCLTV 13

RESULT 14
AAR84501
ID AAR84501 standard; peptide; 20 AA.
XX
AC AAR84501;
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus peptide CORE 29 (residues 169-188).
XX
KW Hepatitis C virus; HCV; immunogen; core region; nucleocapsid;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.

```
XX 28-OCT-1994; 94WO-EP03555.
XX
XX 04-NOV-1993; 93EP-0402718.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Deleys R, Leroux-Roels G, Maertens G;
XX WPI; 1995-193822/25.
XX
XX Hepatitis C Virus immunogenic polypeptide contg. a T-cell
XX stimulating epitope - from core, E1, E2 and NS3 regions, useful in
XX production of vaccines, therapeutic agents, etc.
XX
XX Example 4; Page 51; 105pp; English.
XX
XX A series of overlapping peptides (including the present sequence) was
XX synthesised based on sequences in the core, E1 and E2/NS1 regions of
XX hepatitis C virus. The peptides were used as antigens in lympho-
XX proliferative assays to identify the main T-cell epitopes.
XX
XX Sequence 20 AA;
SQ
Query Match 100.0%; Score 46; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLALLSCLTV 10
Db 10 llallscitv 19
RESULT 15
AAB18534
ID AAB18534 standard; Protein; 135 AA.
XX
XX AAB18534;
DT 15-JAN-2001 (first entry)
XX
XX Protein encoded by a novel hepatitis C virus cDNA clone CA216a.
XX
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX viral infectivity; viral replication.
XX
XX Hepatitis C virus.
XX
XX EP1034785-A2.
XX
XX 13-SEP-2000.
XX
XX 16-MAR-1990; 2000EP-0109602.
XX
XX 17-MAR-1989; 89US-0325338.
XX 20-APR-1989; 89US-0341334.
XX 18-MAY-1989; 89US-0355002.
XX 16-MAR-1990; 90EP-0302866.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
XX
XX WPI; 2000-566891/53.
XX N-PSDB; AAA/5290.
XX
XX Novel composition comprising a hepatitis C virus antisense
XX polynucleotide which is complementary to or corresponds to a sense
XX strand of the virus genome, and selectively hybridises to it -
XX
XX Example; Fig 10; 75pp; English.
```

```
CC The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polyprotein. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises
CC an agent which causes viral RNA to be inactive. The composition
CC is used for preventing HCV replication in a system. The present
CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
XX course of the invention.
XX
XX Sequence 135 AA;
SQ
Query Match 100.0%; Score 46; DB 21; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLALLSCLTV 10
Db 66 llallscitv 75
Search completed: August 23, 2002, 10:00:58
Job time: 388 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:50 ; Search time 61.77 Seconds
(without alignments)
3.954 Million cell updates/sec

Title: US-08-854-825-2
Perfect score: 46
Sequence: 1 LLALLSCLTV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	10	1 US-08-214-650-2	Sequence 2, Appl
2	46	100.0	10	5 PCT-US95-02121-47	Sequence 47, Appl
3	46	100.0	10	5 PCT-US95-02121-126	Sequence 126, App
4	46	100.0	135	4 US-08-444-818-91	Sequence 91, Appl
5	46	100.0	149	4 US-08-836-075A-4	Sequence 4, Appl
6	46	100.0	149	4 US-08-836-075A-24	Sequence 24, Appl
7	46	100.0	149	4 US-08-836-075A-30	Sequence 30, Appl
8	46	100.0	149	4 US-08-836-075A-32	Sequence 32, Appl
9	46	100.0	149	4 US-08-836-075A-34	Sequence 34, Appl
10	46	100.0	149	4 US-08-836-075A-38	Sequence 38, Appl
11	46	100.0	166	2 US-08-483-695-3	Sequence 3, Appl
12	46	100.0	166	2 US-08-483-695-31	Sequence 31, Appl
13	46	100.0	166	2 US-08-483-695-33	Sequence 33, Appl
14	46	100.0	166	2 US-07-965-285-3	Sequence 3, Appl
15	46	100.0	166	2 US-07-965-285-31	Sequence 31, Appl
16	46	100.0	166	2 US-07-965-285-33	Sequence 33, Appl
17	46	100.0	166	2 US-08-487-231-3	Sequence 3, Appl
18	46	100.0	166	2 US-08-487-231-31	Sequence 31, Appl
19	46	100.0	166	2 US-08-487-231-33	Sequence 33, Appl
20	46	100.0	166	4 US-09-201-912-3	Sequence 3, Appl
21	46	100.0	166	4 US-09-201-912-31	Sequence 31, Appl
22	46	100.0	166	4 US-09-201-912-33	Sequence 33, Appl
23	46	100.0	185	4 US-08-612-973-24	Sequence 24, Appl
24	46	100.0	185	4 US-08-927-597-24	Sequence 24, Appl
25	46	100.0	191	2 US-08-290-665A-155	Sequence 155, App
26	46	100.0	191	2 US-08-290-665A-156	Sequence 156, App
27	46	100.0	191	2 US-08-290-665A-157	Sequence 157, App

28	46	100.0	191	2	US-08-290-665A-158	Sequence 158, App
29	46	100.0	191	2	US-08-290-665A-159	Sequence 159, App
30	46	100.0	191	2	US-08-290-665A-160	Sequence 160, App
31	46	100.0	191	2	US-08-290-665A-170	Sequence 170, App
32	46	100.0	191	2	US-08-290-665A-191	Sequence 191, App
33	46	100.0	191	2	US-08-290-665A-192	Sequence 192, App
34	46	100.0	191	2	US-08-290-665A-195	Sequence 195, App
35	46	100.0	191	2	US-08-290-665A-196	Sequence 196, App
36	46	100.0	191	2	US-08-290-665A-197	Sequence 197, App
37	46	100.0	191	4	US-08-380-160-3	Sequence 3, Appli
38	46	100.0	191	5	PCT-US95-10398-155	Sequence 155, App
39	46	100.0	191	5	PCT-US95-10398-156	Sequence 156, App
40	46	100.0	191	5	PCT-US95-10398-157	Sequence 157, App
41	46	100.0	191	5	PCT-US95-10398-158	Sequence 158, App
42	46	100.0	191	5	PCT-US95-10398-159	Sequence 159, App
43	46	100.0	191	5	PCT-US95-10398-160	Sequence 160, App
44	46	100.0	191	5	PCT-US95-10398-170	Sequence 170, App
45	46	100.0	191	5	PCT-US95-10398-191	Sequence 191, App

ALIGNMENTS

RESULT 1
US-08-214-650-2
; Sequence 2, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silvert, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-214-650-2

Query Match 100.0%; Score 46; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLALLSCLTV 10
|||||||

```

; SOFTWARE: PatentIn Release #1.0, Version #1.2.10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02121
; FILING DATE: 16-FEB-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,484
; FILING DATE: 16-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/935,811
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,491
; FILING DATE: 27-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,682
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,568
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14137-26-4PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 126:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PCT-US95-02121-126

Query Match 100.0%; Score 46; DB 5;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0;

QY 1 LLALLSCLTV 10
   |||||
Db 1 LLALLSCLTV 10

RESULT 4
US-08-444-818-91
; Sequence 91, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANV Diagnostics and Vaccin
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590

```


; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-91

Query Match 100.0%; Score 46; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.44; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 1 LLALLSCLTV 10
Db 66 LLALLSCLTV 75

RESULT 5
US-08-836-075A-4
; Sequence 4, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 21 Oct 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 149 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-836-075A-4

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 19 LLALLSCLTV 28

RESULT 6
US-08-836-075A-24
; Sequence 24, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 149 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-836-075A-24

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 19 LLALLSCLTV 28

RESULT 7
US-08-836-075A-30
; Sequence 30, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT

```
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF INVENTIONS: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 149 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-30

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 19 LLALLSCLTV 28

RESULT 8
US-08-836-075A-32
; Sequence 32, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF INVENTIONS: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 149 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-32

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 19 LLALLSCLTV 28

RESULT 9
US-08-836-075A-34
; Sequence 34, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF INVENTIONS: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
```

REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-34

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
| | | | | | | | | |
DB 19 LLALLSCLTV 28

RESULT 10
US-08-836-075A-38
Sequence 38, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836, 075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-38

Query Match 100.0%; Score 46; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
| | | | | | | | | |

DB 19 LLALLSCLTV 28

RESULT 11
US-08-483-695-3
Sequence 3, Application US/08483695
Patent No. 5866139
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Krensdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentLin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,695
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/965,285
FILING DATE: 18-MAR-1993
APPLICATION NUMBER: FR 91 06 882
FILING DATE: 06-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-00000
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-483-695-3

Query Match 100.0%; Score 46; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
| | | | | | | | | |
DB 21 LLALLSCLTV 30

RESULT 12
US-08-483-695-31
Sequence 31, Application US/08483695
Patent No. 5866139
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Krensdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications

; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-483-695-31

Query Match 100.0%; Score 46; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
| | | | | | | | | |
Db 21 LLALLSCLTV 30

RESULT 13
US-08-483-695-33
; Sequence 33, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-483-695-33

Query Match 100.0%; Score 46; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
| | | | | | | | | |
Db 21 LLALLSCLTV 30

RESULT 14
US-07-965-285-3
; Sequence 3, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-965-285-3

Query Match 100.0%; Score 46; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
| | | | | | | |
Db 21 LLALLSCLTV 30

RESULT 15
US-07-965-285-31
; Sequence 31, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brecht, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-965-285-31

Query Match 100.0%; Score 46; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
| | | | | | | |
Db 21 LLALLSCLTV 30

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:42 ; Search time 78.82 Seconds
(without alignments)
12.191 Million cell updates/sec

Title: US-08-854-825-2
Perfect score: 46
Sequence: 1 LLALLSCLTV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	100.0	189	2 S32740	polyprotein - hepa
2	46	100.0	513	2 A4150	structural protein
3	46	100.0	640	2 J01584	genome polyprotein
4	46	100.0	782	2 S18032	genome polyprotein
5	46	100.0	782	2 S19875	genome polyprotein
6	46	100.0	3010	1 S18030	genome polyprotein
7	46	100.0	3011	1 GNMVC3	genome polyprotein
8	46	100.0	3011	1 GNMVCH	genome polyprotein
9	46	100.0	3011	1 S40770	genome polyprotein
10	45	97.8	369	2 S21471	genome polyprotein
11	45	97.8	415	2 PC4407	envelope protein -
12	45	97.8	441	2 S12707	genome polyprotein
13	45	97.8	513	2 PC1284	genome polyprotein
14	45	97.8	782	2 S19876	genome polyprotein
15	45	97.8	782	2 S18031	genome polyprotein
16	45	97.8	787	2 PN0677	hypothetical prote
17	45	97.8	3010	1 GNMVCJ	genome polyprotein
18	45	97.8	3010	1 A45573	genome polyprotein
19	45	97.8	3010	1 GNMVTV	genome polyprotein
20	44	95.7	876	2 PC2219	polypeptide - hepa
21	43	93.5	874	2 J00883	genome polyprotein
22	43	93.5	3033	1 GNMVJ8	genome polyprotein
23	42	91.3	550	2 JH0711	genome polyprotein
24	42	91.3	3010	1 GNMVTC	genome polyprotein
25	40	87.0	494	2 I38967	cytochrome P450 -
26	40	87.0	494	2 I38965	cytochrome P450 -
27	40	87.0	494	2 C34271	genome polyprotein
28	40	87.0	874	2 J00881	genome polyprotein
29	40	87.0	3033	1 J01303	genome polyprotein

30	38	82.6	423	2 T18784	hypothetical prote
31	38	82.6	494	2 A47494	cytochrome P450 2A
32	38	82.6	494	2 B47494	cytochrome P450 2A
33	38	82.6	525	2 T20662	hypothetical prote
34	37	80.4	82	2 T09628	hypothetical prote
35	37	80.4	494	1 O4HUA6	coumarin 7-hydroxy
36	36	78.3	322	2 JN0265	genome polyprotein
37	36	78.3	497	2 I38966	cytochrome P450 -
38	36	78.3	3014	1 JC5620	genome polyprotein
39	35	76.1	308	2 T24732	hypothetical prote
40	35	76.1	335	1 KHB08	cathepsin B (EC 3.
41	35	76.1	597	2 C69283	hypothetical prote
42	34	73.9	199	2 T45543	hypothetical prote
43	34	73.9	387	2 F95069	transmembrane prot
44	34	73.9	409	2 T47026	hypothetical prote
45	34	73.9	409	2 AG0235	probable sugar tra

ALIGNMENTS

RESULT 1

S32740

polyprotein - hepatitis C virus (isolate Russian) (fragment)

N:Contains: capsid protein C; envelope protein M

C:Species: hepatitis C virus

C>Date: 19-Mar-1997 #sequence_revision 05-Dec-1998 #text_change 17-Nov-2000

C:Accession: S32740

R:Vassiliev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.

submitted to the EMBL Data Library, April 1993

A:Description: Evidence of new HCV variant of European isolate in Russia.

A:Reference number: S32740

A:Accession: S32740

A:Molecule type: genomic RNA

A:Residues: 1-189 <VAS>

C:Cross-references: EMBL:X71407

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; envelope protein; polyprotein

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-189/Product: envelope protein M #status predicted <EPM>

Query Match 100.0%; Score 46; DB 2; Length 189;

Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10

|||||

Db 178 LLALLSCLTV 187

RESULT 2

A4150

structural protein - hepatitis C virus

C:Species: hepatitis C virus

C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000

C:Accession: A4150

R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley,

Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992

A:Title: Interaction of immune sera with synthetic peptides corresponding to the stru

A:Reference number: A44150; MUID:92228749

A:Accession: A44150

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: genomic RNA

A:Residues: 1-513 <CHI>

C:Superfamily: hepatitis C virus genome polyprotein

Query Match

Best Local Similarity 100.0%; Score 46; DB 2; Length 513;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10

```

|||||||
Db 178 LLALLSCLTV 187

RESULT 3
JQ1584
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
A:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core
A:Reference number: JQ1584; MUID:92300349
A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUN>
A:Cross-references: GB:X84079; NID:g643119; PIDN:CAA58888.1; PID:g643120
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro
F:1-191/Product: core protein C #status predicted <CPC>
F:192-389/Product: envelope protein E1 #status predicted <BE1>
F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 100.0%; Score 46; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||||
Db 178 LLALLSCLTV 187

RESULT 4
S18032
genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61594
A:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: core protein #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 100.0%; Score 46; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||||
Db 178 LLALLSCLTV 187

RESULT 5
S19875
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)

```

```

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK3
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19875
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus
A:Reference number: S18029
A:Accession: S19875
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
A:Experimental source: isolate JK3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 100.0%; Score 46; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||||
Db 178 LLALLSCLTV 187

RESULT 6
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Variety: isolate JK1
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
A:Accession: S18030; S33570; A48332; S18029
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A:Experimental source: isolate JK1 from an individual
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat
A:Reference number: A48332; MUID:93119270
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-763,'T',763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TCG for residue 3
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPN>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

```


F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 100.0%; Score 46; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
 | | | | | | | | | |
 Db 178 LLALLSCLTV 187

RESULT 7

GNWVC3

N:Contains: genome polyprotein - hepatitis C virus (strain HCV-1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
 C:Accession: A39166; P00403; P00404
 R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A:Title: Genetic organization and diversity of the hepatitis C virus.
 A:Reference number: A39166; MUID:91172826

A:Accession: A39166
 A:Molecule type: mRNA
 A:Residues: 1-3011 <CHO>
 A:Cross-references: GB:M62331; NID:g329873; PIDN:AAA45676.1; PID:g329874
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: P00393; MUID:92268871
 A:Accession: P00403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DDBJ:D10128
 A:Experimental source: Isolates E-b16

A:Accession: P00404
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: Isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 46; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
 | | | | | | | | | |
 Db 178 LLALLSCLTV 187

RESULT 8

GNWVCH

genome polyprotein - hepatitis C virus (strain H)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992

A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp

A:Reference number: A41546; MUID:92052256
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 100.0%; Score 46; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
 | | | | | | | | | |
 Db 178 LLALLSCLTV 187

RESULT 9

S40770

genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PC1285

R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992

A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OKA>
 A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116
 A:Accession: PC1285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NMA>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 46; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 178 LLALLSCLTV 187
|||||

RESULT 10
S21471
genome polyprotein - hepatitis C virus (fragment)
N:Contains: capsid protein; envelope protein
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
C:Accession: S21471
R:Mogam, W.K.
submitted to the EMBL Data Library, April 1992
A:Reference number: S21471
A:Accession: S21471
A:Molecule type: genomic RNA
A:Residues: 1-369 <MOG>
A:Cross-references: EMBL:X65924; NID:g59466; PIDN:CAA46717.1; PID:g59467
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

Query Match 97.8%; Score 45; DB 2; Length 369;
Best Local Similarity 90.0%; Pred. No. 1.4;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 178 LLALLSCLTI 187
|||||

RESULT 11
PC4407
envelope protein - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 10-Nov-1997 #sequence_revision 23-Jan-1998 #text_change 17-Nov-2000
C:Accession: PC4407
R:Li, G.; Yao, J.; Peng, W.
Chinese J. Virol. 13, 24-32, 1997
A:Title: Sequence of genomic region of hepatitis C virus envelope proteins from a Guangd
A:Reference number: PC4407
A:Accession: PC4407
A:Molecule type: genomic RNA
A:Residues: 1-415 <LIA>
A:Note: the authors translated the codon ATA for residues 93 and 249 as Met
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 97.8%; Score 45; DB 2; Length 415;
Best Local Similarity 90.0%; Pred. No. 1.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||

Db 13 LLALLSCLTI 22

RESULT 12
S12707
genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: S12707
R:Takeuchi, K.; Kubo, Y.; Boonmar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.
Nucleic Acids Res. 18, 4626, 1990
A:Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus geno
A:Reference number: S12707; MUID:90356432
A:Accession: S12707
A:Molecule type: genomic RNA
A:Residues: 1-441 <TAK>
A:Cross-references: EMBL:D00574; NID:g221656; PIDN:BAA00452.1; PID:g221657
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 97.8%; Score 45; DB 2; Length 441;
Best Local Similarity 90.0%; Pred. No. 1.7;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 178 LLALLSCLTI 187
|||||

RESULT 13
PC1284
genome polyprotein - hepatitis C virus (isolate HC-J4) (fragment)
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1284
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116
A:Accession: PC1284
A:Molecule type: genomic RNA
A:Residues: 1-513 <OKA>
A:Cross-references: GB:D00832; NID:g221513; PIDN:BAA00706.1; PID:g221514
C:Superfamily: hepatitis C virus genome polyprotein

Query Match 97.8%; Score 45; DB 2; Length 513;
Best Local Similarity 90.0%; Pred. No. 1.9;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 178 LLALLSCLTI 187
|||||

RESULT 14
S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK5
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19876
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus
A:Reference number: S18029
A:Accession: S19876
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487

A:Experimental source: isolate JK5
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 97.8%; Score 45; DB 2; Length 782;
Best Local Similarity 90.0%; Pred. No. 2.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||
Db 178 LLALLSCLTI 187

RESULT 15

S18031

genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK2
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S18031
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029

A:Accession: S18031
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61593

A:Experimental source: isolate JK2

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 97.8%; Score 45; DB 2; Length 782;
Best Local Similarity 90.0%; Pred. No. 2.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||
Db 178 LLALLSCLTI 187

Search completed: August 23, 2002, 09:56:43
Job time: 133 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:17 ; Search time 37 Seconds
(without alignments)
10.465 Million cell updates/sec

Title: US-08-854-825-2
Perfect score: 46
Sequence: 1 LLALLSCLTV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	46	100.0	192	1	POLG_HCVF1
2	46	100.0	309	1	POLG_HCVH7
3	46	100.0	321	1	POLG_HCVH8
4	46	100.0	321	1	POLG_HCVTH
5	46	100.0	3011	1	POLG_HCV1
6	46	100.0	3011	1	POLG_HCVH
7	45	97.8	513	1	POLG_HCVJ2
8	45	97.8	3010	1	POLG_HCVJA
9	45	97.8	3010	1	POLG_HCVJT
10	45	97.8	3010	1	POLG_HCVTW
11	44	95.7	737	1	POLG_HCVJ5
12	43	93.5	737	1	POLG_HCVJ7
13	43	93.5	3033	1	POLG_HCVJ8
14	42	91.3	3010	1	POLG_HCVBK
15	40	87.0	20	1	CPA7_PAPSP
16	40	87.0	494	1	CPA7_HUMAN
17	40	87.0	3033	1	POLG_HCVJ6
18	38	82.6	494	1	CPAA_RABIT
19	38	82.6	494	1	CPAB_RABIT
20	37	80.4	82	1	YBDJ_KLEPN
21	37	80.4	494	1	CPA6_HUMAN
22	36	78.3	494	1	CPAD_HUMAN
23	35.5	77.2	285	1	T13B_HUMAN
24	35	76.1	308	1	VRU3_CAEEL
25	35	76.1	335	1	CATB_BOVIN
26	35	76.1	597	1	Z267_ARCFU
27	34	73.9	520	1	POLG_HCVH4
28	34	73.9	520	1	POLG_HCVHK
29	34	73.9	685	1	PHUB_SALTY
30	33	71.7	128	1	CYB_CROVV
31	33	71.7	323	1	PF27_MOUSE
32	33	71.7	395	1	SOTB_ERWCH
33	33	71.7	472	1	PPB_ESCFE

RESULT 1
POLG_HCVF1
ID POLG_HCVF1 STANDARD: PRT; 192 AA.
AC P27954;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
Major envelope protein E] (Fragment).
DE Hepatitis C virus (isolate EC1) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11107;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
corresponding to the flavivirus envelope and NS1 proteins and the
pestivirus envelope glycoproteins.";
RT Virology 180:842-848(1991).
RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X53135; CAA37295.1; -;
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
transmembrane.
KW NON_TER 1
FT CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 >192 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 20315 MW; 98E488F4C335A84C CRC64;

Query Match 100.0%; Score 46; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||
Db 62 LLALLSCLTV 71

RESULT 2

POLG_HCVH7 STANDARD; PRT; 309 AA.
AC P27955;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT27) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X53133; CAA37293.1; -
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR ProbDom: PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT CHAIN 1
FT NON_TER 1
FT CHAIN <1 63 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 64 255 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 256 >309 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 309 309
SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 309;

Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
|||||
Db 50 LLALLSCLTV 59

RESULT 3

POLG_HCVH8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X53131; CAA37291.1; -
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR ProbDom: PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT CHAIN 1
FT NON_TER 1
FT CHAIN <1 76 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 321;

Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10

Db 62 LLALLSCLTV 71

RESULT 4

POLG_HCVTH STANDARD; PRT; 321 AA.
ID POLG_HCVTH

```
AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (Isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53134; CAA37294.1; -
DR InterPro: IPR002531; HCV_Ns1.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_Ns1; 1.
DR ProDom: PD186062; HCV_Ns1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match 100.08; Score 46; DB 1; Length 321;
Best Local Similarity 100.08; Pred. No. 0.38;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLALLSCLTV 10
    |1111111111
Db 62 LLALLSCLTV 71

RESULT 5
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (Isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53134; CAA37294.1; -
DR InterPro: IPR002531; HCV_Ns1.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_Ns1; 1.
DR ProDom: PD186062; HCV_Ns1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 76 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match 100.08; Score 46; DB 1; Length 321;
Best Local Similarity 100.08; Pred. No. 0.38;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLALLSCLTV 10
    |1111111111
Db 62 LLALLSCLTV 71

RESULT 5
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (Isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=911172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.,
RA "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62321; AAA45676.1; -
DR PIR: A39166; GNWVC3.
DR HSP: P27958; IHEI.
DR MEROPS: S29.001; -
DR MEROPS: U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_Ns1.
DR InterPro: IPR002518; HCV_Ns2.
DR InterPro: IPR004109; HCV_Ns3.
DR InterPro: IPR00745; HCV_Ns4a.
DR InterPro: IPR001490; HCV_Ns4b.
DR InterPro: IPR002868; HCV_Ns5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_Ns1; 1.
DR Pfam: PF01538; HCV_Ns2; 1.
DR Pfam: PF02907; HCV_Ns3; 1.
DR Pfam: PF01006; HCV_Ns4a; 1.
DR Pfam: PF01001; HCV_Ns4b; 1.
DR Pfam: PF01506; HCV_Ns5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR ProDom: PD186062; HCV_Ns1; 1.
DR SMART: SM00492; HELICC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
```

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 191 MATRIX ENVELOPE (POTENTIAL).
 FT CHAIN 383 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 NONSTRUCTURAL PROTEIN NS3 (POTENTIAL).
 FT CHAIN 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 3011 POTENTIAL.
 FT TRANSEM 347 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 ATP (POTENTIAL).
 FT SITE 1316 DECH BOX.
 FT CARBOHYD 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C947FCE5AF9 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred No. 2.6;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALLSCLTV 10
 |||||
 Db 178 LIALLSCLTV 187

RESULT 6
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchausp G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C

RT virus: comparison with American and Japanese isolates.";
 RN Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RL [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hoog Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1; -
 DR PIR; A36814; GNWVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1AIV; 16-FEB-99.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RdRP; 1.
 DR Pfam; PF00271; helicase_C; 1.


```
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 747 809 PROTEIN P7.
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CDD94753 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLALLSCLTV 10
Db 178 LLALLSCLTV 187

RESULT 7
POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
```

```
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D10074; BAA00968.1; -.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CC CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
CC CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
CC CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
CC CHAIN 384 >513 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
CC TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
CC CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
CC NON_TER 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 97.8%; Score 45; DB 1; Length 513;
Best Local Similarity 90.0%; Pred. No. 0.84;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLALLSCLTV 10
Db 178 LLALLSCLTV 187

RESULT 8
POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
```


CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in PI and Ser or Ala in PI'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D11168; RAA01943.1; -.
 CC PIR: A45573; A45573.
 CC HSP: P26663; LJP.
 CC MEROPS: S29.001; -.
 CC
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; HCV_RdRP; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00492; HELTC3; 1.
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
 CC INIT_MET 1 1
 CC
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 729
 CC CHAIN 730 1006
 CC CHAIN 1007 1615
 CC CHAIN 1616 1862
 CC CHAIN 1863 2013
 CC CHAIN 2014 3010
 CC TRANSMEM 347 369
 CC ACT_SITE 1083 1083
 CC ACT_SITE 1107 1107
 CC ACT_SITE 1165 1165
 CC NP_BIND 1230 1237
 CC SITE 1316 1319
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

 Query Match 97.8%; Score 45; DB 1; Length 3010;
 Best Local Similarity 90.0%; Pred. No. 3.9;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 LLALLSCLTV 10
 Db 178 LLALLSCLTV 187

 RESULT 10
 POLG_HCVTVW STANDARD; PRT; 3010 AA.
 ID POLG_HCVTVW
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 FT "The Taiwanese hepatitis C virus genome: sequence determination and
 FT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in PI and Ser or Ala in PI'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC


```

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 95.7%; Score 44; DB 1; Length 737;
Best Local Similarity 90.0%; Pred. No. 1.7;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 178 LLALLSCLTV 187

RESULT 12
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL: D10077; BAA00971.1; -
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC Pfam: PF01543; HCV_capsids; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.

```

```

KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 93.5%; Score 43; DB 1; Length 737;
Best Local Similarity 90.0%; Pred. No. 2.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 178 LLALLSCLTV 187

RESULT 13
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL: D10077; BAA00971.1; -
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC Pfam: PF01543; HCV_capsids; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.

```

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10988; BAA01761.1; -
CC PIR: A40250; GNMVJ8.
CC HSP: P27958; 1HEI.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NSI.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS4c.
CC InterPro: IPR002166; HCV_RDRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NSI; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02307; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS4c; 1.
CC Pfam: PF00998; HCV_RDRP; 1.
CC Pfam: PF00271; helicase_C; 1.
CC ProDom: PD186062; HCV_NSI; 1.
KW Polyprotein: Coated protein; Transferase: RNA-directed RNA polymerase;
KW Core protein; Glycoprotein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT CHAIN 3034 369
FT TRANSMEM 347 369
FT ACT_SITE 1087 1097
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 430
FT CARBOHYD 430 433
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534

FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FDIA CRC64;

Query Match 93.5%; Score 43; DB 1; Length 3033;
Best Local Similarity 90.0%; Pred. No. 8.7;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSLCTV 10
DB 178 LLALLSCVTV 187

RESULT 14
POLG_HCVBK STANDARD; PRT; 3010 AA.
ID AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=911140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parage H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomska Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

Db |:|||||
 8 LVALACTV 17

Search completed: August 23, 2002, 09:55:19
Job time: 49 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:21:58 ; Search time 139.83 Seconds
(without alignments)
12.372 Million cell updates/sec

Title: US-08-854-825-2
Perfect score: 46
Sequence: 1 LLALLSCLTV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	121	12	Q9E2I8 hepatitis c
2	46	100.0	131	12	Q9E2L8 hepatitis c
3	46	100.0	137	12	Q9E2P8 hepatitis c
4	46	100.0	137	12	Q9E2P7 hepatitis c
5	46	100.0	137	12	Q9E2P5 hepatitis c
6	46	100.0	137	12	Q9E2P4 hepatitis c
7	46	100.0	137	12	Q9E2P3 hepatitis c
8	46	100.0	137	12	Q9E2P2 hepatitis c
9	46	100.0	137	12	Q9E2P1 hepatitis c
10	46	100.0	137	12	Q9E2P0 hepatitis c
11	46	100.0	137	12	Q9E2N9 hepatitis c
12	46	100.0	137	12	Q9E2N8 hepatitis c
13	46	100.0	137	12	Q9E2N7 hepatitis c
14	46	100.0	137	12	Q9E2N6 hepatitis c
15	46	100.0	137	12	Q9E2N5 hepatitis c
16	46	100.0	137	12	Q9E2N4 hepatitis c

17	46	100.0	137	12	Q9E2N3 hepatitis c
18	46	100.0	137	12	Q9E2N2 hepatitis c
19	46	100.0	137	12	Q9E2N1 hepatitis c
20	46	100.0	137	12	Q9E2N0 hepatitis c
21	46	100.0	137	12	Q9E2M9 hepatitis c
22	46	100.0	137	12	Q9E2M8 hepatitis c
23	46	100.0	137	12	Q9E2M7 hepatitis c
24	46	100.0	137	12	Q9E2M6 hepatitis c
25	46	100.0	137	12	Q9E2M5 hepatitis c
26	46	100.0	137	12	Q9E2M4 hepatitis c
27	46	100.0	137	12	Q9E2M3 hepatitis c
28	46	100.0	137	12	Q9E2M2 hepatitis c
29	46	100.0	137	12	Q9E2M0 hepatitis c
30	46	100.0	137	12	Q9E2L9 hepatitis c
31	46	100.0	137	12	Q9E2L7 hepatitis c
32	46	100.0	137	12	Q9E2L6 hepatitis c
33	46	100.0	137	12	Q9E2L5 hepatitis c
34	46	100.0	137	12	Q9E2L3 hepatitis c
35	46	100.0	137	12	Q9E2L2 hepatitis c
36	46	100.0	137	12	Q9E2L1 hepatitis c
37	46	100.0	137	12	Q9E2L0 hepatitis c
38	46	100.0	137	12	Q9E2K9 hepatitis c
39	46	100.0	137	12	Q9E2K8 hepatitis c
40	46	100.0	137	12	Q9E2K7 hepatitis c
41	46	100.0	137	12	Q9E2K6 hepatitis c
42	46	100.0	137	12	Q9E2K5 hepatitis c
43	46	100.0	137	12	Q9E2K4 hepatitis c
44	46	100.0	137	12	Q9E2K3 hepatitis c
45	46	100.0	137	12	Q9E2K2 hepatitis c

ALIGNMENTS

RESULT 1
Q9E2I8 PRELIMINARY; PRT; 121 AA.
ID Q9E2I8
AC Q9E2I8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1999;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271877; AAG16213.1; -.
DR InterPro; IPR002519; HCV.env.
DR Pfam; PF01539; HCV.env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 121 AA; 12981 MW; B482A3DCACB5A701 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 121;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

```
RESULT 2
Q9E2L8 ID Q9E2L8 PRELIMINARY; PRT; 131 AA.
AC Q9E2L8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2438;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271847; AAG16183.1; -.
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 131 AA; 14042 MW; BD767FEA1801FOA7 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 131;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 3
Q9E2P8 ID Q9E2P8 PRELIMINARY; PRT; 137 AA.
AC Q9E2P8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2388;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271817; AAG16153.1; -.
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 137 AA; 14780 MW; 44615937EE303988 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 4
Q9E2P7 ID Q9E2P7 PRELIMINARY; PRT; 137 AA.
AC Q9E2P7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3462;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271818; AAG16154.1; -.
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 137 AA; 14734 MW; 8017992B8932346D CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 5
Q9E2P5 ID Q9E2P5 PRELIMINARY; PRT; 137 AA.
AC Q9E2P5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1382;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271820; AAG16156.1; -.
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 137 AA; 14780 MW; 44615937EE303988 CRC64;
```

```
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14745 MW; 12337C518D7EE54C CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
   | | | | | | | |
Db 1 LLALLSCLTV 10

RESULT 6
Q9E2P4
ID Q9E2P4 PRELIMINARY; PRT; 137 AA.
AC Q9E2P4;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2004;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271821; AAG16157.1; -.
DR InterPro: IPR002519; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14770 MW; DAC465A5F2BA559 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
   | | | | | | | |
Db 1 LLALLSCLTV 10

RESULT 7
Q9E2P3
ID Q9E2P3 PRELIMINARY; PRT; 137 AA.
AC Q9E2P3;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2152;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
```

```
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271822; AAG16158.1; -.
DR InterPro: IPR002519; HCV_env; 1.
DR Pfam: PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14837 MW; A6167F0F95FA5E9A CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
   | | | | | | | |
Db 1 LLALLSCLTV 10

RESULT 8
Q9E2P2
ID Q9E2P2 PRELIMINARY; PRT; 137 AA.
AC Q9E2P2;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2440;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271823; AAG16159.1; -.
DR InterPro: IPR002519; HCV_env; 1.
DR Pfam: PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14888 MW; 93EE1C864A85C8AA CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
   | | | | | | | |
Db 1 LLALLSCLTV 10

RESULT 9
Q9E2P1
ID Q9E2P1 PRELIMINARY; PRT; 137 AA.
AC Q9E2P1;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3664;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271824; AAG16160.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
DR Coats protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14889 MW; 77A0481634D0EE21 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 10
Q9E2P0 PRELIMINARY; PRT; 137 AA.
AC Q9E2P0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0800;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271825; AAG16161.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
DR Coats protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14764 MW; 267F9F394F9FD3B7 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 11
Q9E2N9 PRELIMINARY; PRT; 137 AA.
AC Q9E2N9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0873;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271826; AAG16162.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
DR Coats protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14810 MW; 108FB0B82EB1BCD7 CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

RESULT 12
Q9E2N8 PRELIMINARY; PRT; 137 AA.
AC Q9E2N8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0883;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AF271827; AAG16163.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
DR Coats protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14865 MW; 096F0AB88699341A CRC64;

Query Match 100.0%; Score 46; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10

```

RESULT 13
Q9E2N7 Q9E2N6 PRELIMINARY; PRT; 137 AA.
AC Q9E2N7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
(FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0922;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271828; AAG16164.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14846 MW; FFB48C447426B43 CRC64;

Query Match
Best Local Similarity 100.0%; Score 46; DB 12; Length 137;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10
|||||
1 137 137

RESULT 14
Q9E2N6 Q9E2N6 PRELIMINARY; PRT; 137 AA.
AC Q9E2N6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
(FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0923;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271829; AAG16165.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14777 MW; 6E7DA1C046CFF313 CRC64;

Query Match
Best Local Similarity 100.0%; Score 46; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10
|||||
1 137 137

RESULT 15
Q9E2N5 Q9E2N5 PRELIMINARY; PRT; 137 AA.
AC Q9E2N5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
(FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1150;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271830; AAG16166.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14760 MW; D2EF463ED439682E CRC64;

Query Match
Best Local Similarity 100.0%; Score 46; DB 12; Length 137;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLALLSCLTV 10
Db 1 LLALLSCLTV 10
|||||
1 137 137

Search completed: August 23, 2002, 10:21:58
Job time: 1498 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:00:58 ; Search time 179.72 Seconds
(without alignments)
6.180 Million cell updates/sec

Title: US-08-854-825-3
Perfect score: 49
Sequence: 1 QLRHHLLV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA2001.DAT.*

Pred. No... is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	10	AA84572	Cytotoxic T-cell e
2	49	100.0	10	AAV10222	T cell epitope/MHC
3	49	100.0	10	AAV10484	HLA Class I motif
4	49	100.0	10	AAW99361	Vaccine related MH
5	49	100.0	20	AAH84421	Hepatitis C virus
6	49	100.0	20	AAH90972	HCV E1 peptide El-
7	49	100.0	44	AAH20769	Peptide 19 based o
8	49	100.0	44	AAH74234	HCV antigenic Env
9	49	100.0	48	AAH20779	Single branch of o
10	49	100.0	76	AAH84420	Hepatitis C virus
11	49	100.0	90	AAH18533	Protein encoded by

12	49	100.0	106	21	AAH18532	Protein encoded by
13	49	100.0	129	17	AAH90935	HCV S2 domain anti
14	49	100.0	135	21	AAH69652	HCV subtype 1a El
15	49	100.0	147	16	AAH84355	Hepatitis C virus
16	49	100.0	166	13	AAH30062	HCV E1 (envelope p
17	49	100.0	166	20	AAH22020	HCV E1 peptide seq
18	49	100.0	166	20	AAH75481	Hepatitis C virus
19	49	100.0	172	16	AAH79215	pHCV415-encoded AP
20	49	100.0	192	16	AAH69641	Hepatitis C virus
21	49	100.0	192	16	AAH69634	Hepatitis C virus
22	49	100.0	192	16	AAH69639	Hepatitis C virus
23	49	100.0	192	16	AAH69640	Hepatitis C virus
24	49	100.0	192	16	AAH69637	Hepatitis C virus
25	49	100.0	192	16	AAH69635	Hepatitis C virus
26	49	100.0	192	16	AAH69636	Hepatitis C virus
27	49	100.0	192	17	AAH89505	Hepatitis C virus
28	49	100.0	192	17	AAH89506	Hepatitis C virus
29	49	100.0	192	17	AAH89507	Hepatitis C virus
30	49	100.0	192	17	AAH89508	Hepatitis C virus
31	49	100.0	192	17	AAH89510	Hepatitis C virus
32	49	100.0	192	17	AAH89511	Hepatitis C virus
33	49	100.0	192	17	AAH89512	Hepatitis C virus
34	49	100.0	193	14	AAH33982	HCV-1 El protein.
35	49	100.0	193	14	AAH33985	HCT23 El protein.
36	49	100.0	193	14	AAH33984	Th El protein. Sy
37	49	100.0	193	14	AAH33986	HCT27 El protein.
38	49	100.0	193	14	AAH33983	HCT18 El protein.
39	49	100.0	193	14	AAH33987	HC-J1 El protein.
40	49	100.0	211	17	AAH29234	HCV1 El polypeptid
41	49	100.0	211	20	AAH67614	Hepatitis C virus
42	49	100.0	221	16	AAH79213	pHCV172-encoded AP
43	49	100.0	228	14	AAH40116	HGH-HCV-E1 fusion
44	49	100.0	228	16	AAH79214	pHCV168-encoded pr
45	49	100.0	319	15	AAH45330	Anti-HCV antibody

ALIGNMENTS

RESULT	1
AAH84572	
ID	AAH84572 standard; peptide; 10 AA.
XX	AAH84572;
AC	AC
DT	25-APR-1996 (first entry)
XX	
DE	Cytotoxic T-cell epitope, aa 257-266 of HCV-1 El region.
XX	
KW	Hepatitis C virus; HCV; epitope; vaccine; immunogen.
XX	
OS	Hepatitis C virus.
XX	
PN	W09525122-A1.
XX	
PD	21-SEP-1995.
XX	
PF	16-MAR-1995; 95WO-US03224.
XX	
PR	17-MAR-1994; 94US-0214650.
XX	
PA	(SCRI) SCRIPPS RES INST.
XX	
PI	Cerny A, Chisari FV;
XX	
DR	WPI; 1995-336941/43.
XX	
PT	Novel molecule comprising a cytotoxic T cell epitope - used to
PT	vaccinate against hepatitis C viral infection
XX	
PS	Claim 1; Page 60; 85pp; English.
XX	
CC	AAH84570-616, AAH84885-90 and AAH91054 are all HCV-1 derived peptides

CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
 CC were tested for peptide specific cytotoxic T-cell activity. The
 CC peptides AAR84570-77 were found to have substantial homology with a T-
 CC cell epitope and are useful in vaccines against HCV infection.

XX
 SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 16; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0094;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
 Db 1 qlrrhidlliv 10

RESULT 2
 AAY10222
 ID AAY10222 standard; Peptide; 10 AA.

XX
 AC AAY10222;
 XX
 DT 12-MAY-1999 (first entry)
 XX
 DE T cell epitope/MHC ligand SEQ ID NO:152.
 XX
 KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX
 OS Synthetic.

OS Hepatitis C virus.

XX
 PN WO9902183-A2.

XX
 PD 21-JAN-1999.

XX
 PF 10-JUL-1998; 98WO-US14289.

XX
 PR 10-DEC-1997; 97US-0988320.

XX
 PR 10-JUL-1997; 97CA-2209815.

XX
 PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX
 PI Kuendig TM, Simard JJJ;

XX
 DR WPI; 1999-120514/10.

XX
 PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX
 PS Disclosure; Page 29; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX
 SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0094;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
 Db 1 qlrrhidlliv 10

RESULT 3
 AAY10484
 ID AAY10484 standard; Peptide; 10 AA.

XX
 AC AAY10484;

XX
 DT 12-MAY-1999 (first entry)

XX
 DE HLA Class I motif peptide SEQ ID NO:414.

XX
 KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX
 OS Synthetic.

OS Hepatitis C virus.

XX
 PN WO9902183-A2.

XX
 PD 21-JAN-1999.

XX
 PF 10-JUL-1998; 98WO-US14289.

XX
 PR 10-DEC-1997; 97US-0988320.

XX
 PR 10-JUL-1997; 97CA-2209815.

XX
 PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX
 PI Kuendig TM, Simard JJJ;

XX
 DR WPI; 1999-120514/10.

XX
 PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX
 PS Disclosure; Page 42; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX
 SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0094;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHIDLIV 10
 Db 1 qlrrhldllv 10

RESULT 4
 AAM99361
 ID AAM99361 standard; Peptide; 10 AA.
 XX
 AC AAM99361;
 XX
 DT 07-DEC-2001 (first entry)
 XX
 DE Vaccine related MHC ligand peptide SEQ ID NO:464.
 XX
 KW Glutamic acid; glutamine; vaccine; major histocompatibility complex;
 KW MHC; immunomodulator; antiallergic; endocrine; neuroprotectant;
 KW virucidal; bactericidal; antiparasitic; fungicidal; cytostatic;
 KW medicine; pharmaceutical; immune disorder; immune deficiency;
 KW autoimmune; hypersensitivity; allergy; graft rejection; infection;
 KW hormonal disorder; central nervous system disease; cancer; melanoma;
 KW anti-melanoma vaccine; human immunodeficiency virus.
 XX
 OS Hepatitis C virus.
 XX
 PN WO200170772-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 22-MAR-2001; 2001WO-FR00872.
 XX
 PR 23-MAR-2000; 2000FR-0003711.
 XX
 PA (FABR) FABRE MEDICAMENT SA PIERRE.
 XX
 PI Klinguer-Hamour C, Corvaia N, Beck A, Goetsch L;
 XX WPI; 2001-611470/70.
 XX
 XX Stabilized pharmaceutical containing N-terminal glutamic acid or
 PT glutamine, useful e.g. in anti-melanoma vaccines, is an addition salt
 PT with strong acid -
 XX
 PS Claim 9; Page 110; 149pp; French.
 XX
 CC The present invention describes a pharmaceutical compound (I) that
 CC contains an N-terminal glutamic acid (Glu) or glutamine (Gln) residue
 CC in the form of an addition salt with a strong, physiologically
 CC acceptable acid (II). Also described are: (a) a pharmaceutical
 CC composition containing at least one (I); (b) a vaccine containing
 CC at least one (I) where this is a major histocompatibility complex (MHC)
 CC ligand (Ia); (c) a method for in vitro diagnosis of diseases associated
 CC with the presence of (Ia); (d) a kit for method (c) that includes a (Ia);
 CC and (e) a process for preparing (I). (I) has immunomodulator, endocrine,
 CC antiallergic, neuroprotectant, virucidal, bactericidal, antiparasitic,
 CC fungicidal and cytostatic activities. (I) are useful, in human or
 CC veterinary medicine, in pharmaceutical compositions (for treating immune
 CC disorders, e.g. immune deficiency, autoimmune states, hypersensitivity,
 CC allergy, graft rejection, infection, hormonal disorders and central
 CC nervous system diseases), also, where (I) is a MHC ligand (Ia), in
 CC vaccines for treatment or prevention of: (i) viral, bacterial, parasitic
 CC or fungal infections; or (ii) of cancers. A particular application is in
 CC anti-melanoma vaccines. (I) are also useful for in vitro diagnosis of
 CC diseases associated with interactions between MHC and (I), e.g. melanoma
 CC and human immunodeficiency virus infection. AAM98898 to AAM99392
 CC represent peptides which can be used in pharmaceutical compounds from
 CC the present invention.
 XX
 SQ Sequence 10 AA;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHIDLIV 10
 Db 1 qlrrhldllv 10

RESULT 5
 AAR84421
 ID AAR84421 standard; peptide; 20 AA.
 XX
 AC AAR84421;
 XX
 DT 06-JAN-1997 (first entry)
 XX
 DE Hepatitis C virus E1 region (253-272) peptide.
 XX
 KW Hepatitis C virus; HCV; immunogen; E1 region; immunodominant;
 KW T cell epitope; vaccine.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9512677-A2.
 XX
 PD 11-MAY-1995.
 XX
 PF 28-OCT-1994; 94WO-EP03555.
 XX
 PR 04-NOV-1993; 93EP-0402718.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Deleys R. Leroux-Roels G, Maertens G;
 XX WPI; 1995-193822/25.
 DR
 PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
 PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
 PT production of vaccines, therapeutic agents, etc.
 XX
 PS Claim 26; Page 70; 105pp; English.
 XX
 CC Polypeptides comprising 8-104 (pref. 8-68) amino acids from the HCV
 CC E1 region sequence spanning positions 253-332 and
 CC containing a T-cell stimulating epitope are used in HCV immunogenic
 CC compositions. The present sequence is a specifically claimed example
 CC of such a T-cell epitope-containing peptide.
 XX
 SQ Sequence 20 AA;

Query Match 100.0%; Score 49; DB 16; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHIDLIV 10
 Db 5 qlrrhldllv 14

RESULT 6
 AAR90972
 ID AAR90972 standard; peptide; 20 AA.
 XX
 AC AAR90972;
 XX
 DT 25-SEP-1996 (first entry)
 XX
 DE HCV E1 peptide E1-43 for competition studies.
 XX
 KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
 KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

OS Synthetic.
 PN WO9604385-A2.
 XX
 PD 15-FEB-1996.
 XX
 XX 31-JUL-1995; 95WO-EP03031.
 PF
 XX 29-JUL-1994; 94EP-0870132.
 PR
 XX (INNO-) INNOGENETICS NV.
 PA
 PI Bosman F, Buyse M, De Martynoff G, Maertens G;
 XX
 DR WPI; 1996-129401/13.
 XX
 XX Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
 PT proteins - in presence of di:sulphide bond cleavage agent, to
 PT produce proteins suitable for direct use in vaccines or diagnostic
 PT assays of HCV
 XX
 XX Example 7; Page 66; 146pp; English.
 PS
 XX AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
 CC virus (HCV) E1 and E2 peptides used in competition studies. This
 CC sequence represents a synthetic E1 peptide, and corresponds to residues
 CC 253-272 of the E1 protein sequence. These sequences are useful for in
 CC vitro monitoring of HCV disease, or prognosis of the response to
 CC interferon treatment of patients suffering from HCV infection. These
 CC sequences compete with the proteins produced by AAR12704-R12709 and
 CC AAR12961-R12974, which are included in vectors for the production of
 CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
 CC isolated and purified by carrying out a disulphide bond cleavage, or a
 CC reduction step with a disulphide bond cleavage agent, after lysis of
 CC recombinant host cells. The constructs containing the purified HCV
 CC envelope proteins can be used for vaccinating humans against HCV, for in
 CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
 CC for detecting one or more serological types of HCV present in a
 CC biological sample. The constructs can also be immobilised on a solid
 CC substrate and incorporated into a reversed phase hybridisation assay for
 CC determining the presence or the genotype of HCV. The new purification
 CC method preserves the conformation of the recombinantly expressed E1, E2
 CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
 CC using this method are more reactive with human sera than those isolated
 CC by known techniques.
 XX
 XX Sequence 20 AA;
 SQ
 Query Match 100.0%; Score 49; DB 17; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLRHHDLIV 10
 Db 5 qrrrhdliv 14
 |||||
 RESULT 7
 AAR20769
 ID AAR20769 standard; Protein; 44 AA.
 XX
 AC AAR20769;
 XX
 DT 05-MAY-1992 (first entry)
 XX
 DE Peptide 19 based on immunoreactive region of Hepatitis C virus.
 XX
 KW Non-A, non-B hepatitis virus; non-structural protein; vaccine.
 XX
 OS Synthetic.
 XX
 PN EP468527-A.

XX 29-JAN-1992.
 PD
 XX 26-JUL-1991; 91EP-0112620.
 PF
 XX 24-JUN-1991; 91US-0719819.
 PR
 XX 26-JUL-1990; 90US-0558799.
 PR
 XX 07-FEB-1991; 91US-0651735.
 PR
 XX 11-MAR-1991; 91US-0667275.
 PR
 XX (UNBI-) UTD BIOMEDICAL INC.
 PA
 XX Chang YW, Hosein B;
 PI
 XX WPI; 1992-034279/05.
 DR
 XX New synthetic peptide specific for HCV antibodies - for detection
 PT of HCV or NANSHV e.g. by enzyme-linked immunosorbent assay and is
 PT immunogen for preparation of vaccines
 XX
 PS Claim 1; Page 90; 98pp; English.
 XX
 CC This peptide is one of 19 specifically claimed antigens based on the
 CC immunoreactive regions of the envelope protein and non-structural
 CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
 CC invention also concerns analogues, segments, mixtures, conjugates
 CC and polymers of these peptides. The C-terminal amino acid may be
 CC amidated. See AAR20751-R20782.
 CC
 XX Sequence 44 AA;
 SQ
 Query Match 100.0%; Score 49; DB 13; Length 44;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLRHHDLIV 10
 Db 29 qrrrhdliv 38
 |||||
 RESULT 8
 AAR74234
 ID AAR74234 standard; Peptide; 44 AA.
 XX
 AC AAR74234;
 XX
 DT 24-DEC-1995 (first entry)
 XX
 DE HCV antigenic Env peptide SSAL.
 XX
 KW Structured antigenic peptide library; SSAL; vaccine; diagnostic;
 KW therapeutic; hepatitis C virus; HCV.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 2 /note= "V12;E4"
 FT
 FT Misc-difference 3 /note= "R12;K2;N2"
 FT
 FT Misc-difference 4 /note= "E12;V2;D2"
 FT
 FT Misc-difference 5 /note= "G9;S2;N4;D1"
 FT
 FT Misc-difference 6 /note= "N14;G2"
 FT
 FT Misc-difference 7 /note= "A4;S5;F2;T4;V1"
 FT
 FT Misc-difference 8 /note= "S14;L2"
 FT
 FT Misc-difference 9 /note= "R15;H1"
 FT

FT Misc-difference 12 /note= "V12;I4"
FT Misc-difference 13 /note= "A12;Q2;P2"
FT Misc-difference 14 /note= "M3;V6;L7"
FT Misc-difference 15 /note= "T14;S2"
FT Misc-difference 17 /note= "T12;N4"
FT Misc-difference 18 /note= "V9;L7"
FT Misc-difference 20 /note= "T5;A7;V4"
FT Misc-difference 21 /note= "R12;Q2;K2"
FT Misc-difference 22 /note= "D5;N7;Q2;H2"
FT Misc-difference 23 /note= "G5;A2;S2;V2;P2;R2;N1"
FT Misc-difference 24 /note= "K5;S5;T2;G4"
FT Misc-difference 25 /note= "L5;V3;I4;A4"
FT Misc-difference 26 /note= "P12;L4"
FT Misc-difference 27 /note= "A3;T13"
FT Misc-difference 28 /note= "T11;Q2;H1;A1;R1"
FT Misc-difference 29 /note= "Q5;T7;G2;N1;S1"
FT Misc-difference 30 /note= "L10;I6"
FT Misc-difference 32 /note= "R12;T4"
FT Misc-difference 34 /note= "I7;V9"
FT Misc-difference 36 /note= "L12;M4"
FT Misc-difference 37 /note= "L12;I2;V2"
FT Misc-difference 39 /note= "G12;M4"
FT Misc-difference 40 /note= "S7;A8;T1"
FT Misc-difference 42 /note= "T9;A7"
FT Misc-difference 43 /note= "L9;F5;V1;A1"
FT XX
FT XX
PN W09511998-A1.
XX
XX
PD 04-MAY-1995.
XX
XX 26-OCT-1994; 94WO-US12268.
XX
XX 26-OCT-1993; 93US-0143412.
XX
XX (UNBI-) UNITED BIOMEDICAL INC.
XX
XX Hosein B, Kaminsky SM, Koff CW, Kowalski J, Nixon DF;
PI Walfieldam, Wang CY, Ye J, Zamb TJ;
XX
XX WPI; 1995-178890/23.
XX
XX Structured antigenic peptide libraries contain some invariant amino
PT acids - accommodate variations in antigenic structure so are
PT effective against many different strains of e.g. rapidly mutating
PT viruses in vaccines
XX
XX Claim 4; Page 83-84; 216pp; English.
PS
XX

CC In a structured synthetic antigen library, specific amino acids and
CC their frequency of appearance at a variant locus within aligned
CC peptide sequences are defined by the primary sequences of the several
CC variants that make up the alignment used to construct the antigen
CC peptide library. Branched SSALs representing the highly variable
CC yet antigenic regions of the HCV Env and NS1 proteins are given in
CC AAR74233-37 and were used as the key ingredients in a polyvalent
CC vaccine for highly divergent global strains of HCV.
XX
SQ Sequence 44 AA;

Query Match 100.0%; Score 49; DB 16; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLIV 10
Db 29 qlrrhldllv 38

RESULT 9
AAR20779
ID AAR20779 standard; Protein; 48 AA.
XX
AC AAR20779;
XX
DT 05-MAY-1992 (first entry)
XX
DE Single branch of octameric HCV immunogenic peptide #7.
XX
KW Non-A, non-B hepatitis virus; hepatitis C virus;
KW non-structural protein; vaccine; branching poly-L-lysine;
KW octameric branching resin.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 45
FT /label= branch_point
FT /note= "the lysine residue is joined to two 44mer
FT peptides via the alpha and epsilon amino
FT groups"
FT
FT Modified-site 46
FT /label= branch_point
FT /note= "the lysine is joined to Lys(45) via its
FT alpha-amino group and to Lys(47) via its
FT alpha-carboxyl group. The epsilon amino
FT group forms a peptide bond with a lysine
FT group which carries two 44mer peptides"
FT
FT Modified-site 47
FT /label= branch_point
FT /note= "lysine residues are joined to Lys(47) via
FT the alpha and the epsilon amino groups.
FT Lys(47) may be amidated if amino acid 48 is
FT not present"
FT
FT Misc-difference 48
FT /label= Ala, Val, Gly
FT /note= "i.e. is any amino acid containing no side
FT chain functional group or may be absent"
FT
XX
XX EP468527-A.
XX
XX 29-JAN-1992.
XX
XX 26-JUL-1991; 91EP-0112620.
XX
XX 24-JUN-1991; 91US-0719819.
XX 26-JUL-1990; 90US-0558799.
PR 07-FEB-1991; 91US-0651735.
PR 11-MAR-1991; 91US-0667275.
XX
XX (UNBI-) UTD BIOMEDICAL INC.
PA

```

XX PI Chang YW, Hosein B;
XX DR WPI; 1992-034279/05.
XX PT New synthetic peptide specific for HCV antibodies - for detection
XX PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
XX PT immunogen for preparation of vaccines
XX PS Example 19; Page 48; 98pp; English.
XX CC The 4mer peptide component (amino acids 1-44) corresponds to
XX CC peptide 19 (see AAR20769) taken from hepatitis C.
XX CC There are a total of 8 such 4mer peptides in the polymeric
XX CC peptide, each attached to the branching poly-L-lysine resin. Peptide
XX CC immunogens comprising 16, 8, 4 or 2 peptides attached to poly-L-lys
XX CC resin are claimed (see Claim 21). See AAR20751-R20782.
XX SQ Sequence 48 AA;

Query Match 100.0%; Score 49; DB 13; Length 48;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
DB 29 qlrrhldllv 38

RESULT 10
AAR84420
ID AAR84420 standard; peptide; 76 AA.
XX AC AAR84420;
XX DT 06-JAN-1997 (first entry)
XX DE Hepatitis C virus E1 region (253-332) peptide.
XX KW Hepatitis C virus; HCV; immunogen; E1 region; immunodominant;
XX KW T cell epitope; vaccine.
XX OS Hepatitis C virus.
XX PN WO9512677-A2.
XX PD 11-MAY-1995.
XX PF 28-OCT-1994; 94WO-EP03555.
XX PR 04-NOV-1993; 93EP-0402718.
XX PA (INNO-) INNOGENETICS NV.
XX PI Deleys R, Leroux-Roels G, Maertens G;
XX DR WPI; 1995-193822/25.
XX PT Hepatitis C virus immunogenic polypeptide contg. a T-cell
XX PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
XX PT production of vaccines, therapeutic agents, etc.
XX PS Claim 13; Page 65; 105pp; English.
XX CC Polypeptides comprising 8-80 contiguous amino acids from the HCV
XX CC E1 region sequence spanning positions 253-332 (sic) and which
XX CC contain a T-cell stimulating epitope are used in HCV immunogenic
XX CC compositions.
XX SQ Sequence 76 AA;

```

```

Query Match 100.0%; Score 49; DB 16; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
DB 5 qlrrhldllv 14

RESULT 11
AAB18533
ID AAB18533 standard; Protein; 90 AA.
XX AC AAB18533;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a novel hepatitis C virus cDNA clone CA167b.
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX KW viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-0109602.
XX PR 17-MAR-1989; 89US-0325338.
XX PR 20-APR-1989; 89US-0341334.
XX PR 18-MAY-1989; 89US-0355002.
XX PR 16-MAR-1990; 90EP-0302866.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX DR WPI; 2000-566891/53.
XX DR N-PSDB; AAR75289.
XX PT Novel composition comprising a hepatitis C virus antisense
XX PT polynucleotide which is complementary to or corresponds to a sense
XX PT strand of the virus genome, and selectively hybridises to it -
XX PS Example; Fig 9; 75pp; English.
XX CC The specification describes a pharmaceutical composition which
XX CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
XX CC HCV is characterized by a positive stranded RNA genome which has
XX CC 40% homology at the polypeptide level to a HCV polyprotein. The
XX CC antisense polynucleotide binds to cellular polynucleotides which
XX CC enhance and/or are required for viral infectivity, replicative
XX CC ability or chronicity. The antisense polynucleotides may also be
XX CC designed to bind with high specificity, to be of increased stability,
XX CC to be stable and to have low toxicity. The composition also comprises
XX CC an agent which causes viral RNA to be inactive. The composition
XX CC is used for preventing HCV replication in a system. The present
XX CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
XX SQ course of the invention.
XX SQ Sequence 90 AA;

Query Match 100.0%; Score 49; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
DB 61 qlrrhldllv 70

```

```

RESULT 12
AAB18532
ID AAB18532 standard; Protein; 106 AA.
XX AC AAB18532;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a novel hepatitis C virus cDNA clone CA156e.
XX DE Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX KW viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-0109602.
XX PR 17-MAR-1989; 89US-0325338.
XX PR 20-APR-1989; 89US-0341334.
XX PR 18-MAY-1989; 89US-0355002.
XX PR 16-MAR-1990; 90EP-0302866.
XX (CHIR ) CHIRON CORP.
XX PA Houghton M, Choo Q, Kuo G;
XX PI WPI; 2000-566891/53.
XX DR N-PSDB; AAA75288.
XX Novel composition comprising a hepatitis C virus antisense
XX polynucleotide which is complementary to or corresponds to a sense
XX strand of the virus genome, and selectively hybridises to it -
XX Example; Fig 8; 75pp; English.
XX The specification describes a pharmaceutical composition which
XX comprises a hepatitis C virus (HCV) antisense polynucleotide. The
XX HCV is characterized by a positive stranded RNA genome which has
XX 40% homology at the polypeptide level to a HCV polyprotein. The
XX antisense polynucleotide binds to cellular polynucleotides which
XX enhance and/or are required for viral infectivity, replicative
XX ability or chronicity. The antisense polynucleotides may also be
XX designed to bind with high specificity, to be of increased stability,
XX to be stable and to have low toxicity. The composition also comprises
XX an agent which causes viral RNA to be inactive. The composition
XX is used for preventing HCV replication in a system. The present
XX sequence is encoded by a novel HCV cDNA sequence, which is used in the
XX course of the invention.
XX Sequence 106 AA;

Query Match 100.0%; Score 49; DB 21; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHHIDLIV 10
DB 20 qlrrhidllv 29

RESULT 13
AAR90935
ID AAR90935 standard; Protein; 129 AA.
XX AC AAR90935;
XX DT 15-MAY-1996 (first entry)

```

```

XX HCV S2 domain antigen.
XX DE Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;
XX KW diagnosis; antibodies.
XX OS Hepatitis C virus.
XX PN EP693687-A1.
XX PD 24-JAN-1996.
XX PF 03-APR-1991; 91EP-0114016.
XX PR 04-APR-1990; 90US-0504352.
XX (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX WPI; 1996-117956/13.
XX Combinations of synthetic Hepatitis C Virus antigens - provide more
XX effective diagnosis of Non-A, Non-B Hepatitis
XX Claim 7; Fig 1(A-Y); 53pp; English.
XX The combination comprises an HCV antigen from the C domain (pref.
XX C22 - AAR90936) and at least one HCV antigen from the NS3 (pref. C33C
XX - AAR90932), NS4 (pref. C100 - AAR90933), S (pref. S2 - AAR90935) or NS5
XX (AAR90934) domain.
XX The antigens may in the form of a fusion protein, a simple physical
XX mixture, or the individual antigens commonly bound to a solid matrix.
XX They are pref. prepd. by recombinant DNA techniques (primers are
XX given in AAR12711-712716), but can be synthesised or isolated from
XX HCV using affinity chromatography.
XX Sequence 129 AA;

Query Match 100.0%; Score 49; DB 17; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHHIDLIV 10
DB 58 qlrrhidllv 67

RESULT 14
AAY69652
ID AAY69652 standard; Protein; 135 AA.
XX AC AAY69652;
XX DT 08-MAY-2000 (first entry)
XX HCV subtype 1a E1 protein fragment ("Ton").
XX Envelope protein; oligomeric particle; HCV infection; immunogenic;
XX cellular immunity; humoral immunity; E1 protein; subtype 1a;
XX antigen; antibody; vaccine; treatment; prophylaxis; detection.
XX Hepatitis C virus subtype 1a.
XX WO967285-A1.
XX 29-DEC-1999.
XX 23-JUN-1999; 99WO-EP04342.
XX 24-JUN-1998; 98EP-0870142.
XX 22-FEB-1999; 99EP-0870033.

```

```

XX (INNO-) INNOGENETICS NV.
XX Depla E, Maertens G, Bosman A, Van Wijnendaele F;
XX WPI; 2000-147201/13.
XX
XX Novel HCV envelope protein particles used for vaccination against HCV
XX infection -
XX Example 5; Page 51; 105pp; English.
XX
XX The invention relates to a novel oligomeric particle comprising
XX hepatitis C virus (HCV) envelope proteins, with a diameter of 1-100 nm.
XX The invention also encompasses a purified HCV envelope protein
XX (preferably E1 or E1s), specific antibodies generated against the
XX oligomeric particles, or a single HCV envelope protein, and methods for
XX detection of HCV antigens and antibodies. The HCV envelope oligomeric
XX particles are highly immunogenic when presented to chronic HCV carriers,
XX stimulating both cellular and humoral responses, and may additionally
XX comprise a T-cell stimulating antigen such as core protein, E1, E2, NS2,
XX NS3, NS4A, NS4B, NS5A or NS5B. The oligomeric particle of the
XX invention may be used in vaccine compositions against HCV and for
XX inducing immunity against HCV in chronic HCV carriers, especially prior
XX to, simultaneously with, or after any other therapy. They may be used
XX for inducing immunity against HCV in HCV-infected individuals prior to
XX or after liver transplantation or after presumed infection and also for
XX prophylactically inducing immunity against HCV. Antibodies raised
XX against the oligomeric particles are used to detect HCV antigens, and to
XX treat or prevent HCV infection. The oligomeric particles are used to
XX detect HCV antibodies in a sample, and to detect HCV related T cell
XX responses. Prior art methods for HCV vaccination have been unsuccessful.
XX Prophylactic vaccination has also only been shown to be effective
XX against a homologous strain of HCV. The present invention provides
XX methods for successful vaccination against HCV. Sequences
XX AAY69652-Y69654 represent HCV E1 protein fragments referred to in an
XX exemplification of the present invention. AAY69652 represents an E1
XX fragment from HCV subtype 1a and AAY69653-Y60654 represent E1 fragments
XX from two different HCV subtype 1b strains.
XX Sequence 135 AA;

Query Match 100.0%; Score 49; DB 21; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 66 qlrrhldllv 75

RESULT 15
AAR84355
ID AAR84355 standard; protein; 147 AA.
XX
XX AAR84355;
XX
XX 06-JAN-1997 (first entry)
XX
XX Hepatitis C virus E1 region (243-392) immunogenic fragment.
XX
XX Hepatitis C virus; HCV; immunogen; E1 region; immunodominant;
XX T cell epitope; vaccine.
XX
XX Hepatitis C virus.
XX
XX WO9512677-A2.
XX
XX 11-MAY-1995.
XX
XX 28-OCT-1994; 94WO-EP03555.
XX

```

```

PR 04-NOV-1993; 93EP-0402718.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Deleys R, Leroux-Roels G, Maertens G;
XX WPI; 1995-193822/25.
XX
XX Hepatitis C virus immunogenic polypeptide contg. a T-cell
XX stimulating epitope - from core, E1, E2 and NS3 regions, useful in
XX production of vaccines, therapeutic agents, etc.
XX
XX Claim 3; Page 60; 105pp; English.
XX
XX The present polypeptide from the HCV E1 region (spanning positions
XX 243-392) and variants derived from another type of HCV, which
XX contain a T-cell stimulating epitope, can be used in HCV immunogenic
XX compositions.
XX Sequence 147 AA;

Query Match 100.0%; Score 49; DB 16; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 16 qlrrhldllv 25

Search completed: August 23, 2002, 10:00:59
Job time: 389 sec

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:51 ; Search time 61.77 Seconds
(without alignments)
3.954 Million cell updates/sec

Title: US-08-854-825-3

Perfect score: 49

Sequence: 1 QLRRHIDLIV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfilesi.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	10	1	US-08-214-650-3
2	49	100.0	10	1	US-08-214-650-3
3	49	100.0	20	4	US-08-612-973-63
4	49	100.0	20	4	US-08-927-597-63
5	49	100.0	35	1	US-08-927-597-63
6	49	100.0	44	1	US-08-262-037-19
7	49	100.0	44	1	US-08-262-037-129
8	49	100.0	44	1	US-08-262-037-130
9	49	100.0	90	4	US-08-444-818-87
10	49	100.0	106	4	US-08-444-818-85
11	49	100.0	166	2	US-08-483-695-31
12	49	100.0	166	2	US-08-483-695-31
13	49	100.0	166	2	US-08-483-695-33
14	49	100.0	166	2	US-07-965-285-3
15	49	100.0	166	2	US-07-965-285-31
16	49	100.0	166	2	US-07-965-285-33
17	49	100.0	166	2	US-08-487-231-3
18	49	100.0	166	2	US-08-487-231-31
19	49	100.0	166	2	US-08-487-231-33
20	49	100.0	166	4	US-09-201-912-3
21	49	100.0	166	4	US-09-201-912-31
22	49	100.0	166	4	US-09-201-912-33
23	49	100.0	172	1	US-08-188-281B-5
24	49	100.0	172	5	PCT-US94-07280-5
25	49	100.0	172	5	PCT-US95-01087-5
26	49	100.0	192	1	US-08-086-428B-52
27	49	100.0	192	1	US-08-086-428B-53

28	49	100.0	192	1	US-08-086-428B-54	Sequence 54, Appl
29	49	100.0	192	1	US-08-086-428B-55	Sequence 55, Appl
30	49	100.0	192	1	US-08-086-428B-57	Sequence 57, Appl
31	49	100.0	192	1	US-08-086-428B-58	Sequence 58, Appl
32	49	100.0	192	1	US-08-086-428B-59	Sequence 59, Appl
33	49	100.0	192	1	US-08-440-103-37	Sequence 37, Appl
34	49	100.0	192	1	US-08-440-103-38	Sequence 38, Appl
35	49	100.0	192	1	US-08-440-103-40	Sequence 40, Appl
36	49	100.0	192	1	US-08-440-103-41	Sequence 41, Appl
37	49	100.0	192	1	US-08-440-103-44	Sequence 44, Appl
38	49	100.0	192	1	US-08-440-103-45	Sequence 45, Appl
39	49	100.0	192	1	US-08-440-542-37	Sequence 37, Appl
40	49	100.0	192	1	US-08-440-542-38	Sequence 38, Appl
41	49	100.0	192	1	US-08-440-542-40	Sequence 40, Appl
42	49	100.0	192	1	US-08-440-542-41	Sequence 41, Appl
43	49	100.0	192	1	US-08-440-542-44	Sequence 44, Appl
44	49	100.0	192	1	US-08-440-542-45	Sequence 45, Appl
45	49	100.0	192	1	US-08-231-368-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-3
; Sequence 3, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: leydig, Voigt & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-214-650-3

Query Match 100.0%; Score 49; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 QLRRHIDLIV 10
| | | | | | | | | |

Db 1 QLRHIDLIV 10

RESULT 2

US-08-214-650-55

Sequence 55, Application US/08214650

Patent No. 5709995

GENERAL INFORMATION:

APPLICANT: Chisari, Francis V.

APPLICANT: Cerny, Andreas

TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T

TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:

ADDRESSEE: Leydig, Voit & Mayer

STREET: Two Prudential Plaza, Suite 4900

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/214,650

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Silver, Donald J.

REGISTRATION NUMBER: 37552

REFERENCE/DOCKET NUMBER: 61230

TELEPHONE: (312) 616-5600

TELEFAX: (312) 616-5700

TELEX: 25-3533

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-214-650-55

Query Match 100.0%; Score 49; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0035;

Matches 10; Conservative 0; Mismatches 0; Indels 0;

Qy 1 QLRHIDLIV 10

Db 1 QLRHIDLIV 10

RESULT 3

US-08-612-973-63

Sequence 63, Application US/08612973

Patent No. 6150134

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

APPLICANT: BOSMAN, FONS

APPLICANT: DE MARTYNOFF, GUY

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

Query Match 100.0%; Score 49; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0035;

Matches 10; Conservative 0; Mismatches 0; Indels 0;

Qy 1 QLRHIDLIV 10

Db 1 QLRHIDLIV 10

RESULT 4

US-08-927-597-63

Sequence 63, Application US/08927597

Patent No. 6245503

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

APPLICANT: BOSMAN, FONS

APPLICANT: DE MARTYNOFF, GUY

APPLICANT: BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/927,597

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-612-973-63

Query Match 100.0%; Score 49; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.0072;

Matches 10; Conservative 0; Mismatches 0; Indels 0;

Qy 1 QLRHIDLIV 10

Db 5 QLRHIDLIV 14

TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-927-597-63

Query Match 100.0%; Score 49; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLIV 10
Db 5 QLRRHIDLIV 14

RESULT 5

US-08-262-037-81
Sequence 81, Application US/08262037
Patent No. 5747239

GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.

CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 81:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-81

Query Match 100.0%; Score 49; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLIV 10
Db 20 QLRRHIDLIV 29

RESULT 6

US-08-262-037-19

Sequence 19, Application US/08262037

Patent No. 5747239

GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein

TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVE.

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819

FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275

FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735

FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799

APPLICATION NUMBER: 07/510,153

FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 44 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-19

Query Match 100.0%; Score 49; DB 1; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLIV 10
Db 29 QLRRHIDLIV 38

```

RESULT 7
US-08-262-037-129
; Sequence 129, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 129:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-129

```

```

Query Match 100.0%; Score 49; DB 1; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLVL 10
Db 29 QLRRHIDLVL 38

RESULT 8
US-08-262-037-130
; Sequence 130, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

```

```

; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-130

Query Match 100.0%; Score 49; DB 1; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLVL 10
Db 29 QLRRHIDLVL 38

RESULT 9
US-08-444-818-87
; Sequence 87, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/444,818
;; FILING DATE:
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,590
;; FILING DATE: 14-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Harbin, Alisa A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0110.002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (508)359-3876
;; TELEFAX: (508)359-3885
;; INFORMATION FOR SEQ ID NO: 87:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 90 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-444-818-87

Query Match 100.0%; Score 49; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHHIDLLV 10
Db 61 QLRHHIDLLV 70

RESULT 10
US-08-444-818-85
;; Sequence 85, Application US/08444818
;; Patent No. 6150087
;; GENERAL INFORMATION:
;; APPLICANT: Chien, David Y.
;; APPLICANT: Rutter, William J.
;; TITLE OF INVENTION: NANBY Diagnostics and Vaccines
;; NUMBER OF SEQUENCES: 777
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608-2916
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/444,818
;; FILING DATE:
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,590
;; FILING DATE: 14-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Harbin, Alisa A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0110.002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (508)359-3876
;; TELEFAX: (508)359-3885
;; INFORMATION FOR SEQ ID NO: 85:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 106 amino acids
;; TYPE: amino acid

;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-444-818-85

Query Match 100.0%; Score 49; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHHIDLLV 10
Db 20 QLRHHIDLLV 29

RESULT 11
US-08-483-695-3
;; Sequence 3, Application US/08483695
;; Patent No. 5866139
;; GENERAL INFORMATION:
;; APPLICANT: Brechot, Christian
;; APPLICANT: Kremsdorf, Dina
;; APPLICANT: Porchon, Colette
;; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
;; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
;; NUMBER OF SEQUENCES: 46
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;; ADDRESSEE: Dunner
;; STREET: 1300 I Street, N.W.
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA: US/08/483,695
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/965,285
;; FILING DATE: 18-MAR-1993
;; APPLICATION NUMBER: FR 91 06 882
;; FILING DATE: 06-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-00000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 166 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-483-695-3

Query Match 100.0%; Score 49; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRHHIDLLV 10
Db 100 QLRHHIDLLV 109

```
RESULT 12
US-08-483-695-31
; Sequence 31, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-695-31

Query Match 100.0%; Score 49; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 100 QLRRHIDLIV 109

RESULT 13
US-08-483-695-31
; Sequence 33, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
```

```
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-695-33

Query Match 100.0%; Score 49; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 100 QLRRHIDLIV 109

RESULT 14
US-07-965-285-3
; Sequence 3, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-965-285-3

Query Match      100.0%; Score 49; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
   |||||
Db 100 QLRRHIDLIV 109

RESULT 15
US-07-965-285-31
; Sequence 31, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Krensdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: peptide
US-07-965-285-31

Query Match      100.0%; Score 49; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
   |||||
Db 100 QLRRHIDLIV 109

Search completed: August 23, 2002, 09:57:51
Job time: 201 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:43 ; Search time 78.82 Seconds
(without alignments)
12.191 Million cell updates/sec

Title: US-08-854-825-3
Perfect score: 49
Sequence: 1 QLRRHIDLVL 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	513	2 A44150	structural protein
2	49	100.0	640	2 JQ1584	genome polypotein
3	49	100.0	3011	1 GNVVC3	genome polypotein
4	49	100.0	3011	1 GNVVCH	genome polypotein
5	49	100.0	3011	1 S40770	genome polypotein
6	46	93.9	322	2 JN0265	genome polypotein
7	43	87.8	441	2 S12707	genome polypotein
8	42	85.7	550	2 JH0711	genome polypotein
9	42	85.7	787	2 PN0677	hypothetical prote
10	41	83.7	177	2 S32749	genome polypotein
11	41	83.7	177	2 S25123	genome polypotein
12	41	83.7	177	2 S32741	genome polypotein
13	41	83.7	177	2 S32743	genome polypotein
14	41	83.7	177	2 S32745	genome polypotein
15	41	83.7	177	2 S32746	genome polypotein
16	41	83.7	315	2 PS0164	envelope glycoprot
17	41	83.7	315	2 PN0011	envelope glycoprot
18	41	83.7	315	2 PS0165	envelope glycoprot
19	41	83.7	369	2 S21471	genome polypotein
20	41	83.7	415	2 PC4407	envelope protein -
21	41	83.7	513	2 PC1284	genome polypotein
22	41	83.7	520	2 JQ1925	polypotein - hepa
23	41	83.7	523	2 JQ1926	polypotein - hepa
24	41	83.7	782	2 S19876	genome polypotein
25	41	83.7	782	2 S18031	genome polypotein
26	41	83.7	782	2 S18032	genome polypotein
27	41	83.7	782	2 S19875	genome polypotein
28	41	83.7	3010	1 GNVVTC	genome polypotein
29	41	83.7	3010	1 GNVVCJ	genome polypotein

30	41	83.7	3010	1 A45573	genome polypotein
31	41	83.7	3010	1 S18030	genome polypotein
32	41	83.7	3010	1 GNVWTM	genome polypotein
33	39	79.6	180	2 PC1303	genome polypotein
34	39	79.6	428	2 T26316	hypothetical prote
35	39	79.6	471	2 G72518	probable glycine d
36	37	75.5	512	2 A11241	B. subtilis ygpp h
37	37	75.5	1000	2 D87244	conserved hypotet
38	36	73.5	127	2 B84172	hypothetical prote
39	36	73.5	259	2 E86288	hypothetical prote
40	36	73.5	389	2 A83229	hypothetical prote
41	35	71.4	180	2 PC1304	genome polypotein
42	35	71.4	180	2 PC1305	genome polypotein
43	35	71.4	254	2 S72759	hypothetical prote
44	35	71.4	351	2 H72328	cell division prot
45	35	71.4	411	2 PC2060	genome polypotein

ALIGNMENTS

RESULT 1
A44150
structural protein - hepatitis C virus
C:Species: hepatitis C virus
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
C:Accession: A44150
R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
A:Title: Interaction of immune sera with synthetic peptides corresponding to the stru
A:Reference number: A44150; MUID:92228749
A:Accession: A44150
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 1-513 <CHI>
C:Superfamily: hepatitis C virus genome polypotein
Query Match 100.0%; Score 49; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QLRRHIDLVL 10
|||||||
Db 257 QLRRHIDLVL 266
RESULT 2
JQ1584
genome polypotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural p
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative c
A:Reference number: JQ1584; MUID:92300349
A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUM>
C:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120
C:Superfamily: hepatitis C virus genome polypotein
C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; poly
F:1-191/Product: core protein C #status predicted <CPC>
F:192-389/Product: envelope protein E1 #status predicted <EEL>
F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status prediacte
F:196.209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (c

Query Match 100.0%; Score 49; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
| | | | |
Db 257 QLRRHIDLIV 266

RESULT 3
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DBJ:D10128
A:Experimental source: Isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: Isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructu
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 49; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
| | | | |
Db 257 QLRRHIDLIV 266

RESULT 4
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992

A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MUID:92052256
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 100.0%; Score 49; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
| | | | |
Db 257 QLRRHIDLIV 266

RESULT 5
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: Isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 49; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLVL 10
:|||||

Db 257 QLRRHIDLVL 266

RESULT 6

genome polyprotein - hepatitis C virus (isolate GM2) (fragments)
N:Contains: amino end of envelope protein M; carboxyl end of capsid protein C; fragment
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 17-Nov-2000
C:Accession: JN0265
R:Fuchs, K.; Morf, M.; Schreier, E.; Zachoval, R.; Deinhardt, F.; Roggendorf, M.
Gene 103, 163-169, 1991
A:Title: Characterization of nucleotide sequences from European hepatitis C virus isolates
A:Reference number: JN0265; MUID:91365241
A:Accession: JN0265
A:Molecule type: genomic RNA
A:Residues: 1-322 <FUC>
A:Cross-references: GB:M61717; GB:M61718
A:Note: the authors translated the codon ACA for residue 198 as Tyr
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein; transmembrane protein
F:1-109/Product: capsid protein C (fragment) #status predicted <COR>
F:110-178/Product: envelope protein M (fragment) #status predicted <EPM>
F:163-178/Domain: transmembrane #status predicted <TM1>
F:179-322/Product: major envelope protein E (fragment) #status predicted <ENV>
F:253-269/Domain: transmembrane #status predicted <TM2>
F:191.216/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.9%; Score 46; DB 2; Length 322;
Best Local Similarity 90.0%; Pred. No. 0.45;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLVL 10
:|||||

Db 239 ELRRHIDLVL 248

RESULT 7

genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: S12707
R:Takeuchi, K.; Kubo, Y.; Boonmar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.;
Nucleic Acids Res. 18, 4626, 1990
A:Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus genome
A:Reference number: S12707; MUID:90356432
A:Accession: S12707
A:Molecule type: genomic RNA
A:Residues: 1-441 <TAK>
A:Cross-references: EMBL:D00574; NID:g221656; PIDN:BAA00452.1; PID:g221657
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 87.8%; Score 43; DB 2; Length 441;
Best Local Similarity 88.9%; Pred. No. 2.3;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLVL 10
:|||||

Db 258 LRRHIDLVL 266

RESULT 8

genome polyprotein - hepatitis C virus (strain PRC1) (fragments)
N:Contains: envelope protein E1; envelope protein E2; nonstructural protein NS1; nons
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JH0711
R:Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.
Gene 114, 245-250, 1992
A:Title: Genomic typing of hepatitis C viruses present in China.
A:Reference number: JH0711; MUID:92290283
A:Accession: JH0711
A:Molecule type: genomic RNA
A:Residues: 1-550 <LIU>
A:Cross-references: GB:M74888; GB:M74889
A:Note: the nucleotide sequence is not complete
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; nonstructural protein; nucleocapsid; poly
F:1-190/Product: nucleocapsid protein C #status predicted <CPC>
F:191-380/Product: envelope protein E1 #status predicted <EPI>
F:381-514/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #statu
F:515-550/Product: nonstructural protein NS5 (fragment) #status predicted <NS5>
F:196.233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 85.7%; Score 42; DB 2; Length 550;
Best Local Similarity 88.9%; Pred. No. 4.4;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLVL 10
:|||||

Db 258 IRRHIDLVL 266

RESULT 9

hypothetical protein 787 - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PN0677
R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.W.; Lee, C.H.; Lee, Y.I.
Biochem. Biophys. Res. Commun. 196, 780-788, 1993
A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of
A:Reference number: PN0677; MUID:94059104
A:Accession: PN0677
A:Molecule type: mRNA
A:Residues: 1-787 <CHO>
A:Cross-references: GB:I20498; NID:g1381031; PIDN:AAB02608.1; PID:g1381032
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; nonstructural protein
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydr

Query Match 85.7%; Score 42; DB 2; Length 787;
Best Local Similarity 88.9%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLVL 10
:|||||

Db 258 IRRHIDLVL 266

RESULT 10

S32749
genome polyprotein - hepatitis C virus (isolate RU-1) (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus

A:Variety: isolate RU-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32749
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
F:1-177/Product: envelope protein E1 #status predicted <MAT>
A:Reference number: S32741
A:Accession: S32749
A:Molecule type: genomic RNA
A:Residues: 1-177 <ROG>
A:Cross-references: EMBL:X72975; NID:g296114; PIDN:CAA51480.1; PID:g296115
A:Experimental source: isolate RU-1
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: envelope protein; polypeptide
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match 83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLIV 10
:||||:||||
DB 62 IRRHVLLIV 70

RESULT 11
genome polypeptide (clone Glob.1) - hepatitis C virus (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S25123
R:Hoehne, M.; Schreier, E.; Fuchs, K.; Wiese, M.; Deinhardt, F.; Roggendorf, M.
submitted to the EMBL Data Library, April 1992
A:Description: Variability of E1 gene region of Hepatitis C virus in patients infected
A:Reference number: S25123
A:Accession: S25123
A:Molecule type: genomic RNA
A:Residues: 1-177 <HOE>
A:Cross-references: EMBL:X67299; NID:g59470; PIDN:CAA47713.1; PID:g59471
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: envelope protein; polypeptide
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match 83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLIV 10
:||||:||||
DB 62 IRRHVLLIV 70

RESULT 12
genome polypeptide - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus
A:Variety: isolate CR-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32741
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
F:1-177/Product: envelope protein E1 #status predicted <MAT>
A:Reference number: S32741
A:Accession: S32741
A:Molecule type: genomic RNA
A:Residues: 1-177 <ROG>
A:Cross-references: EMBL:X72978; NID:g296100; PIDN:CAA51483.1; PID:g296101
C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: envelope protein; polypeptide
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match 83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLIV 10
:||||:||||
DB 62 IRRHVLLIV 70

RESULT 13
genome polypeptide - hepatitis C virus (isolate EG-1) (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus
A:Variety: isolate EG-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32743
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
F:1-177/Product: envelope protein E1 #status predicted <MAT>
A:Reference number: S32741
A:Accession: S32743
A:Molecule type: genomic RNA
A:Residues: 1-177 <ROG>
A:Cross-references: EMBL:X72980; NID:g296104; PIDN:CAA51485.1; PID:g296105
A:Experimental source: isolate EG-1
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: envelope protein; polypeptide
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match 83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLIV 10
:||||:||||
DB 62 IRRHVLLIV 70

RESULT 14
genome polypeptide - hepatitis C virus (isolate EG-2) (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus
A:Variety: isolate EG-2
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32745
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
F:1-177/Product: envelope protein E1 #status predicted <MAT>
A:Reference number: S32741
A:Accession: S32745
A:Molecule type: genomic RNA
A:Residues: 1-177 <ROG>
A:Cross-references: EMBL:X72982; NID:g296108; PIDN:CAA51487.1; PID:g296109
A:Experimental source: isolate EG-2
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: envelope protein; polypeptide
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match 83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLIV 10
:||||:||||
DB 62 IRRHVLLIV 70

```

RESULT 15
S32746
genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)
N:Contains: envelope protein E1
C:Species: hepatitis C virus
A:Variety: isolate HU-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32746
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its s
A:Reference number: S32741
A:Accession: S32746
A:Molecule type: genomic RNA
A:Residues: 1-177 <ROG>
A:Cross-references: EMBL:X72976; NID:g296110; PIDN:CAA51481.1; PID:g296111
A:Experimental source: isolate HU-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-177/Product: envelope protein E1 #status predicted <MAT>

Query Match      83.7%; Score 41; DB 2; Length 177;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRRHIDLLV 10
Db 62 IRRHVDLLV 70

```

Search completed: August 23, 2002, 09:56:44
Job time: 134 sec

THIS PAGE BLANK (USPTO)

Result No.	Query	Score	Match	Length	DB	ID	Description
1	49	100.0	192	1	POLG_HCVI1	P27954	hepatitis c
2	49	100.0	309	1	POLG_HCVH7	P27954	hepatitis c
3	49	100.0	321	1	POLG_HCVH8	P27956	hepatitis c
4	49	100.0	321	1	POLG_HCVHT	P27957	hepatitis c
5	49	100.0	3011	1	POLG_HCV1	P26664	h genome po
6	49	100.0	3011	1	POLG_HCVH	P27958	h genome po
7	41	83.7	513	1	POLG_HCVJ2	P27959	hepatitis c
8	41	83.7	520	1	POLG_HCVH4	O01404	hepatitis c
9	41	83.7	520	1	POLG_HCVHK	O01404	hepatitis c
10	41	83.7	3010	1	POLG_HCVBK	P26663	h genome po
11	41	83.7	3010	1	POLG_HCVJ4	P26663	h genome po
12	41	83.7	3010	1	POLG_HCVJ7	Q00269	h genome po
13	41	83.7	3010	1	POLG_HCVTW	P29846	h genome po
14	36	73.5	259	1	DEPM_ARATH	Q9fv53	arabidopsi
15	35	71.4	273	1	TC2A_CAERB	Q04202	caenorhabdi
16	35	71.4	351	1	FTSZ2_THEMA	O08398	thermotoga
17	35	71.4	478	1	DISR_AGRKH	P30403	agkistrodon
18	34	69.4	403	1	REFC_AGRRH	P05684	agrobacteri
19	34	69.4	425	1	DHE3_AERRE	Q9yc65	aeropyrum p
20	34	69.4	496	1	F7J3_ANASP	Q05070	anabaena sp
21	34	69.4	676	1	RNR_CHLNP	Q92848	chlamydia p
22	34	69.4	978	1	YEOK_SCHPO	O13816	schizosacch
23	33	67.3	142	1	RLI7_CHLMU	Q9Ejn5	chlamydia m
24	33	67.3	196	1	Y4OT_RHISN	P55605	rhizobium s
25	33	67.3	261	1	YRBE_HAEIN	P45030	haemophilus
26	33	67.3	392	1	FTSZ_NEIGO	P72079	neisseria g
27	33	67.3	392	1	FTSZ_NEIMA	Q51130	neisseria m
28	33	67.3	511	1	CP4B_RAT	P15129	nattus norv
29	33	67.3	692	1	RNR_CHLMU	Q9Bk00	chlamydia m
30	33	67.3	694	1	RNR_CHLTR	O84402	chlamydia t
31	33	67.3	737	1	POLG_HCVJ5	P27960	hepatitis c
32	33	67.3	737	1	POLG_HCVJ7	P27961	hepatitis c
33	33	67.3	738	1	ZN84_HUMAN	P51523	homo sapien

```
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QLRRHIDLIV 10
Db 141 QLRRHIDLIV 150

RESULT 2
POLG_HCVH7
ID POLG_HCVH7 STANDARD; PRT; 309 AA.
AC P27955;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCV27) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53133; CAA37293.1; -
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 63 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 64 255 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 256 >309 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 309
SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;
```

```
Query Match 100.0%; Score 49; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QLRRHIDLIV 10
Db 129 QLRRHIDLIV 138
```

```
RESULT 3
POLG_HCVH8
ID POLG_HCVH8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53131; CAA37291.1; -
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 76 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;
```

```
Query Match 100.0%; Score 49; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QLRRHIDLIV 10
Db 141 QLRRHIDLIV 150
```

```
RESULT 4
POLG_HCVTH
ID POLG_HCVTH STANDARD; PRT; 321 AA.
```


AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (fragment).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53134; CA37294.1; -
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 75
FT CHAIN 76 267
FT CHAIN 268 >321
FT CARBOHYD 80 80
FT CARBOHYD 93 93
FT CARBOHYD 118 118
FT CARBOHYD 189 189
FT CARBOHYD 301 301
FT CARBOHYD 307 307
FT CARBOHYD 314 314
FT NON_TER 321
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match 100.0%; Score 49; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHDLV 10
| | | | | | | | | |
Db 141 QLRRHDLV 150

RESULT 5
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=911172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M62321; AAA45676.1; -
DR PIR: A39166; GNMVC3.
DR HSSP: P27958; 1HEI.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

```

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;

Query Match 100.0%; Score 49; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred No. 0.27;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRRHIDLIV 10
Db 257 QLRRHIDLIV 266

RESULT 6
POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RA "Genomic structure of the human prototype strain H of hepatitis C

```

```

RT virus: comparison with American and Japanese isolates."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain."
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding."
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNMVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1ALV; 16-FEB-99.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.

```

ProDom; PD186062; HCV_NSI; 1.
 SMART; SM00492; HELIC3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 191
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 747 809 PROTEIN P7.
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 100.0%; Score 49; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLRRHIDLIV 10
 |||||
 Db 257 QLRRHIDLIV 266
 RESULT 7
 POLG_HCVJ2 STANDARD; PRT; 513 AA.
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus (isolate HC-J2) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";

Virology 188:331-341(1992).
 CC FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10074; BAA00968.1; -.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 >513 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CARBOHYD 196 196 POTENTIAL.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
 Query Match 83.7%; Score 41; DB 1; Length 513;
 Best Local Similarity 77.8%; Pred. No. 1.4;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LRRHIDLIV 10
 :|||:||||
 Db 258 IRRHVDLLV 266
 RESULT 8
 POLG_HCVH4 STANDARD; PRT; 520 AA.
 ID POLG_HCVH4
 AC Q01404;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HCV-476) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31643;

RN SEQUENCE FROM N.A.
 RP MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10688; BAA01530.1; -.
 DR InterPro; IPR002531; HCV_N51.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002519; HCV_core.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_N51; 1.
 DR ProDom; PD186062; HCV_N51; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW REMOVED FROM CAPSID PROTEIN C BY THE
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 83.7%; Score 41; DB 1; Length 520;
 Best Local Similarity 77.8%; Pred. No. 1.4;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LRRHIDLLV 10
 Db 258 IRRHVDLLV 266
 :|||||

RESULT 9
 POLG_HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1) (Fragment).
 OS Hepatitis C virus (Isolate HCV-KF) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_TaxID=31644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10687; BAA01529.1; -.
 DR PIR; JQ1925; JQ1925.
 DR InterPro; IPR002531; HCV_N51.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002519; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_N51; 1.
 DR ProDom; PD186062; HCV_N51; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW REMOVED FROM CAPSID PROTEIN C BY THE
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; ID2BD0A6FF27349B CRC64;

Query Match 83.7%; Score 41; DB 1; Length 520;
 Best Local Similarity 77.8%; Pred. No. 1.4;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LRRHIDLLV 10
 Db 258 IRRHVDLLV 266
 :|||||

RESULT 10
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 ID POLG_HCVBK
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M.,
RA Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M., Seki M.,
RA Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W.,
RA Theologis A., Ecker J.R.:
RT "Arabidopsis cDNA clones."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins (By similarity).
CC -1- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -1- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial (Potential).
CC -1- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF250959; AAC33973.1; ALT_INIT.
CC EMBL: AC007591; AAD39667.1; -.
CC EMBL: AF361861; AAK32873.1; -.
CC HSSP: P27251; IDFF.
CC InterPro: IPR000181; Pep.deformylase.
CC Pfam: PF01327; Pep.deformylase; 1.
CC ProDom: PD003844; Pep.deformylase; 1.
CC Protein biosynthesis; Hydrolase; Iron; Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 259 PEPTIDE DEFORMYLASE.
FT METAL 178 178 IRON (BY SIMILARITY).
FT METAL 220 220 IRON (BY SIMILARITY).
FT ACT_SITE 221 221 BY SIMILARITY.
FT METAL 224 224 IRON (BY SIMILARITY).
SQ SEQUENCE 259 AA; 28913 MW; F524D19C9D21463D CRC64;

Query Match 73.5%; Score 36; DB 1; Length 259;
Best Local Similarity 70.0%; Pred. No. 6.2;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QLRHHIDLLV 10
| | | | |
Db 149 QERRHFDLMV 158

RESULT 15
TC2A_CAEBR TC2A_CAEBR STANDARD; PRT; 273 AA.
AC Q04202;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Transposable element TCB2 transposase.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9122478; PubMed=1851119;
RA Prasad S.S., Harris L.J., Baillie D.L., Rose A.M.;
RT "Evolutionarily conserved regions in Caenorhabditis transposable
RT elements deduced by sequence comparison."
RL Genome 34:6-12(1991).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M64308; AAA28149.1; -.
CC InterPro: IPR002492; Transposase_5.
CC Pfam: PF01498; Transposase_5; 1.
CC ProDom: PD002059; Transposase_5; 1.
CC Transposable element; DNA-binding; DNA recombination; DNA integration;
CC Nuclear protein.
CC SEQUENCE 273 AA; 31916 MW; C11FDBFA02CCE52B CRC64;

Query Match 71.4%; Score 35; DB 1; Length 273;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RRRHIDLL 9
| | | | |
Db 193 RRRHVDLL 199

Search completed: August 23, 2002, 09:55:21
Job time: 51 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:21:58 ; Search time 139.83 Seconds
(without alignments)
12.372 Million cell updates/sec

Title: US-08-854-825-3
Perfect score: 49
Sequence: 1 QLRRHIDLIV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_19.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	95	12 Q68302	Q68302 hepatitis c
2	49	100.0	102	12 Q68300	Q68300 hepatitis c
3	49	100.0	102	12 Q68301	Q68301 hepatitis c
4	49	100.0	125	12 Q68921	Q68921 hepatitis c
5	49	100.0	125	12 Q68924	Q68924 hepatitis c
6	49	100.0	125	12 Q68925	Q68925 hepatitis c
7	49	100.0	125	12 Q68920	Q68920 hepatitis c
8	49	100.0	126	12 Q68922	Q68922 hepatitis c
9	49	100.0	126	12 Q68923	Q68923 hepatitis c
10	49	100.0	126	12 Q68919	Q68919 hepatitis c
11	49	100.0	128	12 Q68168	Q68168 hepatitis c
12	49	100.0	128	12 Q68169	Q68169 hepatitis c
13	49	100.0	128	12 Q68170	Q68170 hepatitis c
14	49	100.0	128	12 Q68186	Q68186 hepatitis c
15	49	100.0	137	12 Q6E2P7	Q6E2P7 hepatitis c
16	49	100.0	142	12 Q81306	Q81306 hepatitis c

17	49	100.0	149	12 Q68732	Q68732 hepatitis c
18	49	100.0	149	12 Q68733	Q68733 hepatitis c
19	49	100.0	149	12 Q68739	Q68739 hepatitis c
20	49	100.0	166	12 Q9PX23	Q9PX23 hepatitis c
21	49	100.0	166	12 Q68418	Q68418 hepatitis c
22	49	100.0	166	12 Q68419	Q68419 hepatitis c
23	49	100.0	166	12 Q68420	Q68420 hepatitis c
24	49	100.0	166	12 Q68421	Q68421 hepatitis c
25	49	100.0	192	12 Q81416	Q81416 hepatitis c
26	49	100.0	192	12 Q81417	Q81417 hepatitis c
27	49	100.0	192	12 Q81418	Q81418 hepatitis c
28	49	100.0	192	12 Q81384	Q81384 hepatitis c
29	49	100.0	192	12 Q81390	Q81390 hepatitis c
30	49	100.0	192	12 Q81397	Q81397 hepatitis c
31	49	100.0	192	12 Q81402	Q81402 hepatitis c
32	49	100.0	295	12 Q98UM1	Q98UM1 hepatitis c
33	49	100.0	309	12 Q68926	Q68926 hepatitis c
34	49	100.0	309	12 Q98UL7	Q98UL7 hepatitis c
35	49	100.0	314	12 Q98UK7	Q98UK7 hepatitis c
36	49	100.0	315	12 Q98UL4	Q98UL4 hepatitis c
37	49	100.0	315	12 Q98UL3	Q98UL3 hepatitis c
38	49	100.0	315	12 Q98UL2	Q98UL2 hepatitis c
39	49	100.0	315	12 Q98UL1	Q98UL1 hepatitis c
40	49	100.0	315	12 Q98UL0	Q98UL0 hepatitis c
41	49	100.0	315	12 Q98UK8	Q98UK8 hepatitis c
42	49	100.0	315	12 Q98UN7	Q98UN7 hepatitis c
43	49	100.0	316	12 Q98UL9	Q98UL9 hepatitis c
44	49	100.0	318	12 Q98UM2	Q98UM2 hepatitis c
45	49	100.0	318	12 Q98UM0	Q98UM0 hepatitis c

ALIGNMENTS

RESULT 1
Q68302 PRELIMINARY; PRT; 95 AA.
AC Q68302;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
GN E1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-BB43;
RA Songsiivilai S., Kanistanon D.;
RT "Identification and characterisation of Thai isolates of hepatitis' C virus."
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: U23396; AAA64862.1; -.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
FT Transmembrane.
FT NON_TER 1 95
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10305 MW; 86270AA69397533A CRC64;

Query Match 100.0%; Score 49; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
| | | | | | | | | |
DB 39 QLRRHIDLIV 48

```

RESULT 2
Q68300
ID Q68300 PRELIMINARY: PRT: 102 AA.
AC Q68300;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
GN E1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-BB4;
RA Songsivilai S., Kanistanon D.;
RT "Identification and characterisation of Thai isolates of hepatitis' C
virus";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; U23394; AAA64860.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11164 MW; 751A979A260D9B2A CRC64;

```

```

Query Match 100.0%; Score 49; DB 12; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QLRRHIDLIV 10
Db 40 QLRRHIDLIV 49

```

```

RESULT 3
Q68301
ID Q68301 PRELIMINARY: PRT: 102 AA.
AC Q68301;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
GN E1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-BB8;
RA Songsivilai S., Kanistanon D.;
RT "Identification and characterisation of Thai isolates of hepatitis' C
virus";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; U23395; AAA64861.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11040 MW; 2C560825E0AD043E CRC64;

```

```

Query Match 100.0%; Score 49; DB 12; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QLRRHIDLIV 10
Db 40 QLRRHIDLIV 49

```

```

RESULT 4
Q68921
ID Q68921 PRELIMINARY: PRT: 125 AA.
AC Q68921;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H47;
RA Argentin C., Guisepetti R., D'ugo E., Rapicetta M.;
RT "In vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84982; CAA59342.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 125 125
SQ SEQUENCE 125 AA; 13555 MW; E5FAAE14A85D04DA CRC64;

```

```

Query Match 100.0%; Score 49; DB 12; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QLRRHIDLIV 10
Db 65 QLRRHIDLIV 74

```

```

RESULT 5
Q68924
ID Q68924 PRELIMINARY: PRT: 125 AA.
AC Q68924;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H78;
RA Argentin C., Guisepetti R., D'ugo E., Rapicetta M.;
RT "In vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

```

CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84985; CAA59345.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1 1
FT NON_TER 125 125
SQ SEQUENCE 125 AA; 13615 MW; 9DE92A29EA525B3D CRC64;

Query Match 100.0%; Score 49; DB 12; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDL 10
| | | | | | | |
Db 65 QLRRHIDL 74

RESULT 6
Q68925 PRELIMINARY; PRT; 125 AA.
AC Q68925;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H79;
RA Argentin C., Guiseppe R., D'ugo E., Rapicetta M.;
RT "in vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84986; CAA59346.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1 1
FT NON_TER 125 125
SQ SEQUENCE 125 AA; 13634 MW; 1FFD690E0265F55C CRC64;

Query Match 100.0%; Score 49; DB 12; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDL 10
| | | | | | | |
Db 65 QLRRHIDL 74

RESULT 7
Q68920 PRELIMINARY; PRT; 125 AA.
AC Q68920;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H406;
RA Argentin C., Guiseppe R., D'ugo E., Rapicetta M.;
RT "in vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84988; CAA59348.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1 1
FT NON_TER 125 125
SQ SEQUENCE 125 AA; 13571 MW; 8682479F31393689 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDL 10
| | | | | | | |
Db 65 QLRRHIDL 74

RESULT 8
Q68922 PRELIMINARY; PRT; 126 AA.
AC Q68922;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H48;
RA Argentin C., Guiseppe R., D'ugo E., Rapicetta M.;
RT "in vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84983; CAA59343.1; -.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1 1
FT NON_TER 126 126
SQ SEQUENCE 126 AA; 13770 MW; EF1E1139C663C932 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 126;
Best Local Similarity 100.0%; Pred. No. 0.065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDL 10
| | | | | | | |
Db 66 QLRRHIDL 75

RESULT 9
Q68923 PRELIMINARY; PRT; 126 AA.
ID Q68923

AC Q68923;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H68;
RA Argentin C., Guisepetti R., D'ugo E., Rapicetta M.;
RT "In vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84984; CAAS9344.1; -;
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1
FT NON_TER 1
FT NON_TER 126 126
SQ SEQUENCE 126 AA; 13540 MW; C9BCE4CE516B94FC CRC64;

Query Match 100.0%; Score 49; DB 12; Length 126;
Best Local Similarity 100.0%; Pred. No. 0.065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 66 QLRRHIDLIV 75
| | | | | | | | | |

RESULT 10
Q68919
ID Q68919 PRELIMINARY; PRT; 126 AA.
AC Q68919;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H338;
RA Argentin C., Guisepetti R., D'ugo E., Rapicetta M.;
RT "In vivo" molecular variability of HCV E1 protein, in equilibrium
RT and disequilibrium systems following an epidemiologically linked
RT infection.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; X84987; CAAS9347.1; -;
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1
FT NON_TER 1
FT NON_TER 126 126
SQ SEQUENCE 126 AA; 13763 MW; 81PCDF0232DC9662 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 126;
Best Local Similarity 100.0%; Pred. No. 0.065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 66 QLRRHIDLIV 75
| | | | | | | | | |

RESULT 11
Q68168
ID Q68168 PRELIMINARY; PRT; 128 AA.
AC Q68168;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1A;
RX MEDLINE=95146953; Pubmed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; U14204; AAC53893.1; -;
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1
FT NON_TER 1
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14041 MW; E3C99FE837321362 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLRRHIDLIV 10
Db 66 QLRRHIDLIV 75
| | | | | | | | | |

RESULT 12
Q68169
ID Q68169 PRELIMINARY; PRT; 128 AA.
AC Q68169;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1A;
RX MEDLINE=95146953; Pubmed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; U14205; AAC53894.1; -;
DR InterPro; IPR002519; HCV_env.

DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 128
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14041 MW; E3C99FEC6321362 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRHHIDLLV 10
| | | | | | | | | |
Db 66 QLRHHIDLLV 75

RESULT 13
Q68170 PRELIMINARY; PRT; 128 AA.
AC Q68170;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
GN E1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1A;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia.";
RL J. Gen. Virol. 76:211-215(1995).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: U14206; AAC53895.1; -;
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 128
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14027 MW; E3C99FE837360362 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRHHIDLLV 10
| | | | | | | | | |
Db 66 QLRHHIDLLV 75

RESULT 14
Q68186 PRELIMINARY; PRT; 128 AA.
AC Q68186;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
GN E1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

Search completed: August 23, 2002, 10:21:58
Job time: 1498 sec

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1A;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia.";
RL J. Gen. Virol. 76:211-215(1995).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: U14222; AAC53911.1; -;
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 128
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 13999 MW; 5EA7EC4702688414 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRHHIDLLV 10
| | | | | | | | | |
Db 66 QLRHHIDLLV 75

RESULT 15
Q9E2P7 PRELIMINARY; PRT; 137 AA.
AC Q9E2P7;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35)]
DE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3462;
RX MEDLINE=20407394; PubMed=10950762;
RA Ray S.C., Arthur R.R., Carella A., Bukh J., Thomas D.L.;
RT "Genetic Epidemiology of Hepatitis C Virus throughout Egypt.";
RL J. Infect. Dis. 182:698-707(2000).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF271818; AAG16154.1; -;
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01539; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14734 MW; 8017992B8932346D CRC64;

Query Match 100.0%; Score 49; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.07;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLRHHIDLLV 10
| | | | | | | | | |
Db 80 QLRHHIDLLV 89

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:00:59 ; Search time 179.72 seconds
(without alignments)
5.562 Million cell updates/sec

Title: US-08-854-825-26

Perfect score: 50

Sequence: 1 LLCPCAGHAV 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 segs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	16	AA84573
2	50	100.0	9	20	AAV10236
3	50	100.0	9	20	AAV10517
4	50	100.0	10	16	AA84599
5	50	100.0	93	10	AAFP0148
6	50	100.0	93	10	AAFP2031
7	50	100.0	93	12	AA14541
8	50	100.0	93	12	AA14352
9	50	100.0	93	16	AA86842
10	50	100.0	93	18	AAW01691
11	50	100.0	93	19	AAW46392

12	50	100.0	93	20	AAW97604	Amino acid sequenc
13	50	100.0	182	21	AA15211	Hepatitis C virus
14	50	100.0	183	17	AAW12963	HCV NS3 protease c
15	50	100.0	183	17	AAW04578	Hepatitis C virus
16	50	100.0	183	18	AAW01641	HCV NS3 protease c
17	50	100.0	183	19	AAW47143	HCV NS3 protease N
18	50	100.0	183	21	AAW57202	HCV NS3 protease.
19	50	100.0	184	21	AAW44727	Hepatitis C virus
20	50	100.0	187	19	AAW56480	Amino acid sequenc
21	50	100.0	189	17	AAW12965	HCV NS3 protease-s
22	50	100.0	189	18	AAW01643	Hepatitis C virus
23	50	100.0	191	21	AAW44728	Hepatitis C virus
24	50	100.0	195	21	AAW15212	Hepatitis C virus
25	50	100.0	195	21	AAW15220	Hepatitis C virus
26	50	100.0	197	21	AAW15221	Hepatitis C virus
27	50	100.0	197	21	AAW15222	Hepatitis C virus
28	50	100.0	197	21	AAW15223	Hepatitis C virus
29	50	100.0	197	21	AAW15224	Hepatitis C virus
30	50	100.0	197	21	AAW15225	Hepatitis C virus
31	50	100.0	197	21	AAW15226	Hepatitis C virus
32	50	100.0	201	17	AAW12966	HCV solubilised NS
33	50	100.0	201	17	AAW04573	HCV NS3 construct
34	50	100.0	201	18	AAW01644	HCV NS3 soluble pr
35	50	100.0	202	16	AAW86548	Hepatitis C virus
36	50	100.0	202	18	AAW14752	HCV NS3 domain pro
37	50	100.0	202	18	AAW01696	HCV NS3 domain pro
38	50	100.0	202	19	AAW46346	Hepatitis C virus
39	50	100.0	202	20	AAW97597	Protein sequence o
40	50	100.0	202	20	AAW97613	HCV protease seque
41	50	100.0	210	17	AAW09241	HCV insoluble NS3
42	50	100.0	210	17	AAW04582	HCV NS3 protease c
43	50	100.0	210	18	AAW01650	HCV NS3 protease c
44	50	100.0	234	17	AAW09240	HCV insoluble NS3
45	50	100.0	234	18	AAW01649	HCV NS3 protease c

ALIGNMENTS

RESULT 1

AA84573

ID AAR84573 standard; peptide: 9 AA.

XX AAR84573;

DT 25-APR-1996 (first entry)

XX Cytotoxic T-cell epitope, aa 1169-1177 of HCV-1 NS3 region.

DE Hepatitis C virus; HCV; epitope; vaccine; immunogen.

KW Hepatitis C virus.

OS Hepatitis C virus.

XX WO9525122-A1.

PN 21-SEP-1995.

XX 16-MAR-1995; 95WO-US03224.

XX 17-MAR-1994; 94US-0214650.

XX (SCRI) SCRIPPS RES INST.

PA Cerny A, Chisari FV;

XX WPI; 1995-336941/43.

DR Novel molecule comprising a cytotoxic T cell epitope - used to

PT vaccinate against hepatitis C viral infection

XX Claim 1; Page 67; 85pp; English.

XX AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides

CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
 CC were tested for peptide specific cytotoxic T-cell activity. The
 CC peptides AAR84570-77 were found to have substantial homology with a T-
 CC cell epitope and are useful in vaccines against HCV infection.
 XX
 SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 16; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
 DB 1 llcpghav 9

RESULT 2
 AAY10236
 ID AAY10236 standard; Peptide: 9 AA.

XX AC AAY10236;
 XX DT 12-MAY-1999 (first entry)
 XX T cell epitope/MHC ligand SEQ ID NO:166.
 XX Cytotoxic T-lymphocyte response; CTL; antigen: lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX OS Synthetic.
 OS Hepatitis C virus.
 XX PN WO9902183-A2.
 XX PD 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-US14289.
 XX PR 10-DEC-1997; 97US-0988320.
 XX PR 10-JUL-1997; 97CA-2209815.

PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX PI Kuendig TM, Simard JLL;
 XX DR WPI; 1999-120514/10.

PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
 of antigen in the lymphatic system of a mammal so as to provide a
 sustained CTL response, used to treat, e.g. AIDS

PS Disclosure; Page 30; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 method comprises: (a) delivering an antigen to the mammal at a level to
 induce an immunological CTL response in the mammal; and (b) maintaining
 the level of the antigen in the mammal's lymphatic system to maintain
 the immunologic CTL response. The method can be used for the delivery of
 e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 gene antigen, or a viral antigen. They can be used for the treatment of
 disease such as cancer, e.g. malignant melanoma or infectious disease,
 e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 to the lymphatic system provides for potent CTL stimulation that takes
 place in the milieu of the lymphoid organ, and it sustains stimulation
 that is necessary to keep CTL active, cytotoxic and recirculating
 through the body. AAY10071 to AAY10639 represent examples of peptide
 antigens given in the present invention.

XX Sequence 9 AA;

Query Match 100.0%; Score 50; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
 DB 1 llcpghav 9

RESULT 3
 AAY10517
 ID AAY10517 standard; Peptide: 9 AA.

XX AC AAY10517;
 XX DT 12-MAY-1999 (first entry)
 XX HLA Class I motif peptide SEQ ID NO:447.
 XX Cytotoxic T-lymphocyte response; CTL; antigen: lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX OS Synthetic.
 OS Hepatitis C virus.
 XX PN WO9902183-A2.

XX PD 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-US14289.

XX PR 10-DEC-1997; 97US-0988320.

XX PR 10-JUL-1997; 97CA-2209815.

PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX PI Kuendig TM, Simard JLL;

XX DR WPI; 1999-120514/10.

PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
 of antigen in the lymphatic system of a mammal so as to provide a
 sustained CTL response, used to treat, e.g. AIDS

PS Disclosure; Page 44; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 method comprises: (a) delivering an antigen to the mammal at a level to
 induce an immunological CTL response in the mammal; and (b) maintaining
 the level of the antigen in the mammal's lymphatic system to maintain
 the immunologic CTL response. The method can be used for the delivery of
 e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 gene antigen, or a viral antigen. They can be used for the treatment of
 disease such as cancer, e.g. malignant melanoma or infectious disease,
 e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 to the lymphatic system provides for potent CTL stimulation that takes
 place in the milieu of the lymphoid organ, and it sustains stimulation
 that is necessary to keep CTL active, cytotoxic and recirculating
 through the body. AAY10071 to AAY10639 represent examples of peptide
 antigens given in the present invention.

XX Sequence 9 AA;

Query Match 100.0%; Score 50; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
|
Db 1 llcpaghav 9

RESULT 4

AAR84599
ID AAR84599 standard; peptide: 10 AA.

XX AC AAR84599;

XX DT 25-APR-1996 (first entry)

XX DE HCV-1 derived peptide tested for usefulness in a HCV vaccine.

XX KW Hepatitis C virus; HCV; epitope; vaccine; immunogen.

XX OS Hepatitis C virus.

XX PN WO9525122-A1.

XX PD 21-SEP-1995.

XX PF 16-MAR-1995; 95WO-US03224.

XX PR 17-MAR-1994; 94US-0214650.

XX PA (SCRI) SCRIPPS RES INST.

XX PI Cerny A, Chisari FV;

XX DR WPI; 1995-336941/43.

XX PT Novel molecule comprising a cytotoxic T cell epitope - used to
PT vaccinate against hepatitis C viral infection

XX PS Example 1; Page 45; 85pp; English.

XX CC AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides
CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
CC were tested for peptide specific cytotoxic T-cell activity. The
CC peptides AAR84570-77 were found to have substantial homology with a T-
CC cell epitope and are useful in vaccines against HCV infection.

XX SQ sequence 10 AA;

Query Match 100.0%; Score 50; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
|
Db 2 llcpaghav 10

RESULT 5

AAP90148
ID AAP90148 standard; protein; 93 AA.

XX AC AAP90148;

XX DT 01-NOV-1989 (first entry)

XX DE Sequence of hepatitis C virus cDNA insert in clone 8h.

XX KW Hepatitis C virus; clone 8h; clone 33c; probe; vaccine.

XX OS Pan troglodytes.

XX FH Key Location/Qualifiers

XX FT Region 71..93

PN GB22125111-A.

XX PD 26-JUL-1989.

XX PF 18-NOV-1988; 88GB-0027024.

XX PR 18-NOV-1987; 87US-0122714.

XX PA (CHIR) CHIRON CORPORATION.

XX PI Houghton M, Choo QL, Kuo G;

XX DR WPI; 1989-215054/30.

XX DR N-PSDB; AAN90317.

XX PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.

XX PS Disclosure; fig 16; 235pp; English.

XX CC The sequence is the peptide encoded by the hepatitis C virus

CC (HCV) cDNA insert in clone 8h (see AAN90317). The polypeptides
CC are used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.

XX CC The region shown overlaps with clone 33c.

XX SQ Sequence 93 AA;

Query Match 100.0%; Score 50; DB 10; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9

|
Db 48 llcpaghav 56

RESULT 6

AAP92031
ID AAP92031 standard; protein; 93 AA.

XX AC AAP92031;

XX DT 02-MAR-1990 (first entry)

XX DE Sequence encoded in the hepatitis C virus (HCV) cDNA insert in clone 8h.

XX KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

XX FT misc_feature 71..93

XX FT /*tag= a

XX PN EP318216-A.

XX PD 31-MAY-1989.

XX PF 18-NOV-1988; 88EP-0310922.

XX PR 14-NOV-1988; 88US-0271450; US-122714.

XX PA (CHIR) CHIRON CORP.

XX PI Houghton M, Choo q-L, Kuo G;

XX DR WPI; 1989-159274/22.

XX DR N-PSDB; AAN92087.

XX PT Purified hepatitis C virus

PT - and associated nucleic acids and polypeptide(s)

PS Claim 13; Figure 16; 139pp; English.

CC It is the sequence encoded in the hepatitis C virus (HCV) cDNA insert in
 CC clone 8h. Tag a - the region of overlap with the HCV antigen encoded in
 CC clone 33c. It is an epitope which could be used as immunoassay reagents
 CC and vaccines and to generate antibodies useful in diagnosis and passive
 CC immunotherapy for HCV infection/non-A, non-B hepatitis.

XX Sequence 93 AA;

Query Match 100.0%; Score 50; DB 10; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.28;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

DB 48 llcpaghav 56

RESULT 7

AAR14541
 ID AAR14541 standard; Protein; 93 AA.

XX AC AAR14541;

XX DT 14-JAN-1992 (first entry)

XX DE Encoded by Hepatitis C Virus protease clone C8h.

XX KW NANBH; non-A, non-B hepatitis; liver disease; fusion protein.
 XX OS Hepatitis C virus.
 XX PN WO9115575-A.

XX PD 17-OCT-1991.

XX PF 04-APR-1991; 91WO-US02210.

XX PR 04-APR-1990; 90US-0505433.

XX PA (CHIR-) CHIRON CORP.

XX PI Houghton M, Choo Q, Kuo G;

XX DR WPI; 1991-325218/44.

XX DR N-PSDB; AAR14299.

XX PT New purified protease - derived from hepatitis C virus, for

XX assay, and designing anti-HCV agents

XX PS Example 3; Fig 4; 74pp; English.

XX Clone C8h was isolated from a HCV cDNA library (ATCC 40394) using
 CC probe C8h (see AAR14309). It was digested with EcoRI and DdeI to
 CC give a 208bp fragment. This fragment, along with fragments from
 CC clone C33c (see AAR14303) and clone C26d (see AAR14298), was cloned into
 CC the EcoRI site of pBR322 and transformed into E.coli HB101 to give
 CC vector C300. A 945bp NaeI/EcoRI fragment from C300 was ligated to a
 CC fragment from C7f+C20c (see AAR14297 and AAR14300). A fragment from this
 CC recombinant vector was eventually used to construct a vector which
 CC encodes amino acids 1-151 of human Superoxide dismutase fused to amino
 CC acids 946-1630 of HCV protease. The vector (cf1SODp600) was
 CC transformed into E.coli DL1210 cells and deposited as ATCC 68275.

XX Sequence 93 AA;

PT Query Match 100.0%; Score 50; DB 12; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.28;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

DB 48 llcpaghav 56

RESULT 9

AAR68542

ID AAR68542 standard; Protein; 93 AA.

XX AC AAR68542;

XX DT 17-AUG-1995 (first entry)

XX DE Hepatitis C virus (HCV) protease clone C8h.

XX KW Hepatitis C virus protease; HCV; clone C8h;

XX viral infectivity inhibition.

PT Query Match 100.0%; Score 50; DB 12; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.28;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

DB 48 llcpaghav 56

RESULT 9

AAR68542

ID AAR68542 standard; Protein; 93 AA.

XX AC AAR68542;

XX DT 17-AUG-1995 (first entry)

XX DE Hepatitis C virus (HCV) protease clone C8h.

XX KW Hepatitis C virus protease; HCV; clone C8h;

XX viral infectivity inhibition.

```

XX OS Hepatitis C virus.
XX PN US5371017-A.
XX PD 06-DEC-1994.
XX PF 04-APR-1990; 90US-0505433.
XX PR 04-APR-1990; 90US-0505433.
XX PR 04-APR-1991; 91US-0680296.
XX PA (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX WPI; 1995-021889/03.
DR N-PSDB; AAQ80170.
XX DNA encoding hepatitis C virus protease - used to produce large
PT amts. of the protease and to develop prods. for inhibition of
PT viral infectivity
XX Example 3; Fig 4; 69pp; English.
XX AAQ80170 encodes AAR68542 hepatitis C virus (HCV) protease clone C8h,
CC using recombinant expression systems large amounts of protease can
CC be produced. The HCV protease can be used in the production of Abs.
CC It can also be used for assaying agents which inhibit protease
CC activity, to identify compounds which inhibit viral infectivity.
XX Sequence 93 AA;

Query Match 100.0%; Score 50; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
Db 48 llcpaghav 56

RESULT 10
AAW01691
ID AAW01691 standard; Protein; 93 AA.
XX AC AAW01691;
XX DT 03-APR-1997 (first entry)
XX DE HCV protease clone C8h.
XX KW HCV; NS3; non-structural domain 3; protease; polypeptide; inhibitor;
XX KW screen; processing; infection; treatment; probe; hepatitis C virus.
XX OS Hepatitis C Virus.
XX PN US5585258-A.
XX PD 17-DEC-1996.
XX PF 04-APR-1990; 90US-0505433.
XX PR 04-APR-1991; 91US-0680296.
XX PR 04-APR-1990; 90US-0505433.
XX PR 06-DEC-1994; 94US-0350884.
XX PA (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX WPI; 1997-051175/05.

```

```

DR N-PSDB; AAT59255.
XX Compsn. contg. hepatitis C virus NS3 domain protease and related
PT fusion proteins - useful for screening specific inhibitors,
PT potential antiviral agents, prepn. of antibodies and for cleaving
PT specific poly:peptide(s)
XX Example 3; Column 65-68; 68pp; English.
XX Compsns. comprising the hepatitis C virus (HCV) NS3 domain protease or
CC its active truncation analogues are claimed. Also new are fusion
CC proteins comprising the protease (or analogues) and, e.g. human
CC superoxide (SOD) or ubiquitin. The protease is essential for polypeptide
CC processing, and thus infectivity, in HCV. The compsns. are used to screen
CC for specific inhibitors (possibly useful as antiviral agents), to
CC generate specific antibodies and to cleave specific polypeptides. HCV
CC cDNA clones (AAT59250-56 encoding AAW01686-92 resp.) were isolated from
CC HCV genomic library using probes AAT59244-49. The clones were used in the
XX preparation of full-length SOD-protease fusion proteins.
XX Sequence 93 AA;

Query Match 100.0%; Score 50; DB 18; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
Db 48 llcpaghav 56

RESULT 11
AAW46392
ID AAW46392 standard; Protein; 93 AA.
XX AC AAW46392;
XX DT 07-MAY-1998 (first entry)
XX DE Amino acid sequence of the Hepatitis c virus (HCV) clone C8h.
XX KW Protease; HCV; NS3 domain; human superoxide dismutase; fusion protein;
XX KW assay; activity; anti-HCV.
XX OS Hepatitis c virus.
XX PN US5712145-A.
XX PD 27-JAN-1998.
XX PF 06-SEP-1996; 96US-0709173.
XX PR 04-APR-1991; 91US-0680296.
XX PR 04-APR-1990; 90US-0505433.
XX PR 06-DEC-1994; 94US-0350884.
XX PR 12-MAY-1995; 95US-0440548.
XX PA (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX WPI; 1998-119986/11.
DR N-PSDB; AAV04986.
XX Recombinant hepatitis C virus protease - useful in screening drugs
PT for activity against hepatitis C virus
XX Disclosure; Fig 4; 68pp; English.
XX The present sequence represents the amino acid sequence of the
CC Hepatitis C virus (HCV) clone C8h. The clone was isolated using
CC hybridisation probe AAV04975. A cDNA fragment encoding protease was

```

CC isolated from the clone, and cloned into an expression vector to
 CC produce a fusion protein with human superoxide dismutase-protease. The
 CC HCV protease is believed to cleave itself from the genomic polyprotein.
 CC In the absence of protease activity, the HCV polyprotein should remain
 CC in its unprocessed form, and thus render the virus non-infectious.
 CC Inhibitors of protease activity should therefore also inhibit viral
 CC infectivity. The protease can therefore be used for assaying compounds
 CC for activity against HCV.

XX Sequence 93 AA;

Query Match 100.0%; Score 50; DB 19; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
 DB 48 llcpaghav 56

RESULT 12
 AAW97604
 ID AAW97604 standard; Protein; 93 AA.

XX AAW97604;

XX 26-MAY-1999 (first entry)

DE Amino acid sequence of HCV protease clone C8h.

XX HCV NS3 protease; truncation analog; HCV control; protease activity;
 KW viral infectivity; inactive non-cleaving protease.

OS Hepatitis C virus.

XX US5885799-A.

PN 23-MAR-1999.

XX 06-SEP-1996; 96US-0709177.

XX 04-APR-1991; 91US-0680296.

PR 04-APR-1990; 90US-0505433.

PR 06-DEC-1994; 94US-0350884.

PR 12-MAY-1995; 95US-0440548.

PR 06-SEP-1996; 96US-0709177.

XX (CHIR) CHIRON CORP.

XX Choo Q, Houghton M, Kuo G;

DR WPI: 1999-228536/19.

DR N-PSDB; AAX26393.

XX Preparation of new Hepatitis C Virus NS3 protease - useful for
 PT screening for compounds which inhibit HCV infectivity

XX Example 3; Fig 4; 7lpp; English.

XX The specification describes a method for making a purified Hepatitis
 CC C virus (HCV) NS3 protease or active truncation analog. If the HCV
 CC protease N-terminal cleavage signal is excluded (so that self-cleavage
 CC is prevented), the HCV protease remains in its unprocessed form, and
 CC renders the virus noninfectious. The protease is therefore useful for
 CC assaying pharmaceutical agents for control of HCV, as compounds which
 CC inhibit protease activity sufficiently will also inhibit viral
 CC infectivity. An inactive non-cleaving protease can be used to screen
 CC for inhibitors. Recombinant expression systems can be utilised to
 CC prepare recombinant HCV which can be used to produce monoclonal
 CC antibodies. The present sequence was isolated in the course of the
 CC invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 50; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
 DB 48 llcpaghav 56

RESULT 13
 AAB15211
 ID AAB15211 standard; protein; 182 AA.

XX AAB15211;

XX 19-DEC-2000 (first entry)

DE Hepatitis C virus NS3 protease.

KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer.

XX Hepatitis C virus.

PN WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

DR WPI: 2000-465976/40.

DR N-PSDB; AAA70344.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C

XX Claim 3; Fig 9; 66pp; English.

XX The present sequence is the Hepatitis C virus (HCV) NS3 protease enzyme.
 CC This protein is essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. NS4A is also needed for this process and inhibitors of the
 CC two proteins should act as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease:inhibitor complexes.

SQ Sequence 182 AA;

Query Match 100.0%; Score 50; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 0.53;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
 DB 144 llcpaghav 152

RESULT 14

AAW12963
ID AAW12963 standard; Protein; 183 AA.

XX AC AAW12963;
XX DT 29-MAR-1997 (first entry)
XX DE HCV NS3 protease catalytic domain.
XX KW HCV; NS3 protease; substrate: nonstructural polyprotein;
XX KW inhibitor; assay; liver disease; hepatocellular carcinoma; tumour.
XX OS Hepatitis C virus.

XX PN WO9635717-A2.
XX PD 14-NOV-1996.
XX PF 09-MAY-1996; 96WO-US06389.
XX PR 12-MAY-1995; 95US-0439747.
XX PA (SCHE) SCHERING CORP.

XX PI Murray MG, Ramanathan L, Zhang R;
XX DR WPI: 1996-518617/51.
XX DR N-PSDB; AAT42386.

XX PT New soluble substrates for hepatitis C virus NS3 protease - are
PT non-structural poly:proteins and are attached to solubilising
PT motifs, useful for determining protease inhibitors

XX PS Example 1; Page 37-39; 70pp; English.

XX CC The catalytic domain (AAW12963) of hepatitis C virus (HCV) NS3
CC protease (NS3p) can be produced in soluble form by attachment of
CC a solubilising peptide (see also AAW12964-66 and AAW09236-37) or
CC hydrophilic tail (AAW42391-92, AAW09245-47), or in insoluble form
CC (AAW09240), pref. with an N-terminal His tag (AAW09241) to facilitate
CC purification. NS3p may also be expressed as a fusion to NS4A
CC cofactor (AAW09242). NS3p constructs can be produced in host cells
CC using vectors contg. the encoding cDNA sequences (see also AAW42386-95)
CC for use with novel NS3p substrates (AAW12957-62) in high throughput
CC assays to identify HCV protease inhibitors of potential therapeutic
CC appln.

XX SQ Sequence 183 AA;

Query Match 100.0%; Score 50; DB 17; Length 183;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPCGHAV 9
Db 143 llcpaghav 151

RESULT 15

AAW04578
ID AAW04578 standard; Protein; 183 AA.

XX AC AAW04578;
XX DT 08-FEB-1997 (first entry)
XX DE Hepatitis C virus NS3 protease catalytic domain.
XX DE Serine protease; NS3; HCV; inclusion body; protein solubilisation;
XX KW refolding; renaturation.
XX OS Hepatitis C virus.

XX WO9635709-A1.

XX PD 14-NOV-1996.
XX PF 09-MAY-1996; 96WO-US06388.
XX PR 13-DEC-1995; 95US-0571643.
XX PR 12-MAY-1995; 95US-0439680.
XX PA (SCHE) SCHERING CORP.

XX PI Ramanathan L, Wendel M;
XX DR WPI: 1996-518613/51.
XX DR N-PSDB; AAT43706.

XX PT Prodn. of soluble, active HCV NS3 protease - from insoluble
PT aggregates produced by bacteria, using denaturing and reducing agent
XX PS Disclosure; Page 22-24; 36pp; English.

XX CC The catalytic domain (AAW04578) of hepatitis C virus (HCV) NS3
CC protease comprises the 183 N-terminal amino acids of the enzyme.
CC NS3 cleaves the non-structural proteins that are necessary for HCV
CC replication and is a target for the development of therapeutic
CC agents. The enzyme is expressed as insoluble aggregates in E.
CC coli (see also AAT43706). Refolded, soluble, active NS3 protease
CC is obt'd. by: solubilising the aggregates in a buffer contg. a
CC denaturant; placing the solubilised protease in acidic pH buffer
CC contg. a reducing agent; removing the denaturant while maintaining
CC acidic pH; and raising the pH, stepwise, to about 7-8. The NS3
CC protease may be expressed with a poly-histidine tag (see also
CC AAW04582) to facilitate purification.

XX SQ Sequence 183 AA;

Query Match 100.0%; Score 50; DB 17; Length 183;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPCGHAV 9
Db 143 llcpaghav 151

Search completed: August 23, 2002, 10:00:59
Job time: 389 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 23, 2002, 09:57:51 ; Search time 61.77 Seconds
(without alignments)
3.559 Million cell updates/sec

Title: US-08-854-825-26
Perfect score: 50
Sequence: 1 LLCPAGHAV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	9	1	US-08-214-650-26
2	50	100.0	10	1	US-08-214-650-25
3	50	100.0	93	1	US-08-350-884-76
4	50	100.0	93	1	US-08-709-173-76
5	50	100.0	93	2	US-08-709-177-76
6	50	100.0	93	4	US-08-444-818-34
7	50	100.0	182	4	US-09-478-479B-1
8	50	100.0	195	4	US-09-478-479B-3
9	50	100.0	195	4	US-09-478-479B-12
10	50	100.0	197	4	US-09-478-479B-14
11	50	100.0	197	4	US-09-478-479B-16
12	50	100.0	197	4	US-09-478-479B-18
13	50	100.0	197	4	US-09-478-479B-20
14	50	100.0	197	4	US-09-478-479B-22
15	50	100.0	197	4	US-09-478-479B-24
16	50	100.0	202	1	US-08-350-884-1
17	50	100.0	202	1	US-08-350-884-65
18	50	100.0	202	1	US-08-709-173-1
19	50	100.0	202	1	US-08-709-173-65
20	50	100.0	202	2	US-08-709-177-1
21	50	100.0	202	2	US-08-709-177-65
22	50	100.0	299	1	US-08-350-884-66
23	50	100.0	299	1	US-08-350-884-68
24	50	100.0	299	1	US-08-709-173-66
25	50	100.0	299	1	US-08-709-173-68
26	50	100.0	299	2	US-08-709-177-66
27	50	100.0	299	2	US-08-709-177-68

28 50 100.0 474 4 US-08-867-611-26 Sequence 26, Appl
29 50 100.0 474 5 PCT-US92-06965A-31 Sequence 31, Appl
30 50 100.0 631 2 US-08-833-678A-1 Sequence 1, Appl
31 50 100.0 631 4 US-09-128-314-2 Sequence 2, Appl
32 50 100.0 631 4 US-08-529-169A-1 Sequence 1, Appl
33 50 100.0 638 4 US-09-288-391-25 Sequence 25, Appl
34 50 100.0 686 1 US-08-350-884-70 Sequence 70, Appl
35 50 100.0 686 1 US-08-709-173-70 Sequence 70, Appl
36 50 100.0 686 2 US-08-709-177-70 Sequence 70, Appl
37 50 100.0 687 1 US-08-188-281B-14 Sequence 14, Appl
38 50 100.0 687 5 PCT-US94-07280-14 Sequence 14, Appl
39 50 100.0 687 5 PCT-US95-01087-14 Sequence 14, Appl
40 50 100.0 829 4 US-08-444-818-69 Sequence 69, Appl
41 50 100.0 841 1 US-08-350-884-86 Sequence 86, Appl
42 50 100.0 841 1 US-08-709-173-86 Sequence 86, Appl
43 50 100.0 841 2 US-08-709-177-86 Sequence 86, Appl
44 50 100.0 1648 1 US-08-188-281B-12 Sequence 12, Appl
45 50 100.0 1648 5 PCT-US94-07280-12 Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-26
; Sequence 26, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerdy, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/214,650
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-214-650-26

Query Match 100.0%; Score 50; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLCPAGHAV 9
| | | | | | | | | |

Db 1 LLCPAGHAV 9

RESULT 2

US-08-214-650-25

; Sequence 25, Application US/08214650

; Patent No. 5709995

; GENERAL INFORMATION:

; APPLICANT: Chisari, Francis V.

; APPLICANT: Cerny, Andreas

; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leydig, Voit & Mayer

; STREET: Two Prudential Plaza, Suite 4900

; CITY: Chicago

; STATE: IL

; COUNTRY: USA

; ZIP: 60601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/214,650

; FILING DATE: 06-SEP-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/680,296

; FILING DATE: 04-APR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: CIOTTI, THOMAS E.

; REGISTRATION NUMBER: 21,013

; REFERENCE/DOCKET NUMBER: 22300-20100.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 76:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-214-650-25

Query Match 100.0%; Score 50; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0076;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Db 2 LLCPAGHAV 10

RESULT 3

US-08-350-884-76

; Sequence 76, Application US/08350884

; Patent No. 5585258

; GENERAL INFORMATION:

; APPLICANT: HOUGHTON, MICHAEL

; APPLICANT: CHOO, QUI LIM

; APPLICANT: KUO, GEORGE

; TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

Query Match 100.0%; Score 50; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0076;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Db 2 LLCPAGHAV 10

RESULT 4

US-08-709-173-76

; Sequence 76, Application US/08709173

; Patent No. 5712145

; GENERAL INFORMATION:

; APPLICANT: HOUGHTON, MICHAEL

; APPLICANT: CHOO, QUI LIM

; APPLICANT: KUO, GEORGE

; TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

Query Match 100.0%; Score 50; DB 1; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.072;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Db 48 LLCPAGHAV 56

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/350,884

; FILING DATE: 06-DEC-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/680,296

; FILING DATE: 04-APR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: CIOTTI, THOMAS E.

; REGISTRATION NUMBER: 21,013

; REFERENCE/DOCKET NUMBER: 22300-20100.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 76:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-350-884-76

Query Match 100.0%; Score 50; DB 1; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.072;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Db 48 LLCPAGHAV 56

RESULT 4

US-08-709-173-76

; Sequence 76, Application US/08709173

; Patent No. 5712145

; GENERAL INFORMATION:

; APPLICANT: HOUGHTON, MICHAEL

; APPLICANT: CHOO, QUI LIM

; APPLICANT: KUO, GEORGE

; TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

Query Match 100.0%; Score 50; DB 1; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.072;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Db 48 LLCPAGHAV 56

; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-709-173-76

Query Match 100.0%; Score 50; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCAGHAV 9
Db 48 LLCAGHAV 56

RESULT 5

US-08-709-177-76
; Sequence 76, Application US/08709177
; Patent No. 5885799
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI LIM
; APPLICANT: KUO, GEORGE
; TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,177
; FILING DATE: 06-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/680,296
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: CIOTTI, THOMAS E.
; REGISTRATION NUMBER: 21,013
; REFERENCE/DOCKET NUMBER: 22300-20100.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-709-177-76

Query Match 100.0%; Score 50; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCAGHAV 9
Db 48 LLCAGHAV 56

RESULT 6

US-08-444-818-34
; Sequence 34, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-34

Query Match 100.0%; Score 50; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCAGHAV 9
Db 48 LLCAGHAV 56

RESULT 7

US-09-478-479B-1
; Sequence 1, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease-Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; CURRENT FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-1

Query Match      100.0%; Score 50; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPAGHAV 9
Db 144 LLCPAGHAV 152

RESULT 8
US-09-478-479B-3
; Sequence 3, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-3

Query Match      100.0%; Score 50; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.15; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPAGHAV 9
Db 156 LLCPAGHAV 164

RESULT 9
US-09-478-479B-12
; Sequence 12, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-12

Query Match      100.0%; Score 50; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.15; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPAGHAV 9
Db 158 LLCPAGHAV 166

RESULT 11
US-09-478-479B-16
; Sequence 16, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; CURRENT FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-16
```

Query Match 100.0%; Score 50; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 158 LLCPAGHAV 166

RESULT 12
US-09-478-479B-18
; Sequence 18, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-18

Query Match 100.0%; Score 50; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 158 LLCPAGHAV 166

RESULT 13
US-09-478-479B-20
; Sequence 20, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-20

Query Match 100.0%; Score 50; DB 4; Length 197;

Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 158 LLCPAGHAV 166

RESULT 14
US-09-478-479B-22
; Sequence 22, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-22

Query Match 100.0%; Score 50; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 158 LLCPAGHAV 166

RESULT 15
US-09-478-479B-24
; Sequence 24, Application US/09478479B
; Patent No. 6333186
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/478,479B
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-478-479B-24

Query Match 100.0%; Score 50; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
| | | | | | | |
Db 158 LLCPAGHAV 166

Search completed: August 23, 2002, 09:57:52
Job time: 202 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:44 ; Search time 78.82 Seconds
(without alignments)
10.972 Million cell updates/sec

Title: US-08-854-825-26

Perfect score: 50

Sequence: 1 LLCPAGHAV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	3011	1 GNVVC3	genome polyprotein
2	47	94.0	3010	1 A45573	genome polyprotein
3	47	94.0	3010	1 S18030	genome polyprotein
4	46	92.0	3011	1 GNVVCH	genome polyprotein
5	46	92.0	3011	1 S40770	genome polyprotein
6	44	88.0	3010	1 GNVVTC	genome polyprotein
7	43	86.0	3010	1 GNVVCJ	genome polyprotein
8	43	86.0	3010	1 GNVVTW	genome polyprotein
9	42	84.0	3033	1 J01303	genome polyprotein
10	39	78.0	3014	1 JC5620	genome polyprotein
11	38	76.0	177	2 G90806	probable polyprote
12	38	76.0	177	2 C85666	probable polyprote
13	38	76.0	177	2 B64847	probable polyprote
14	38	76.0	179	2 AG2780	hypothetical prote
15	36	72.0	621	2 S35092	plakoglobin - mous
16	36	72.0	738	2 S35093	plakoglobin - Afr
17	36	72.0	744	2 A32905	plakoglobin, desmo
18	36	72.0	781	2 S35099	beta-catenin - Afr
19	36	72.0	781	2 A38973	beta-catenin - hum
20	36	72.0	781	2 S35091	beta-catenin - mou
21	36	72.0	817	2 S33793	hypothetical prote
22	36	72.0	843	2 T12689	armadillo segment
23	36	72.0	972	2 A70619	protein kinase ABC c
24	36	72.0	988	1 S35362	protein kinase C (
25	36	72.0	1252	2 S36016	ococyst wall protei
26	36	72.0	1820	2 A55494	latent transformin
27	35	70.0	326	2 T20776	hypothetical prote
28	35	70.0	331	2 G91133	probable collagena
29	35	70.0	331	2 B65106	probable proteinase

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, O.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

J:Chen, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,556,576,623,645,1213,1255,2041,2077

Query Match 100.0% Score 50; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

|||||

Db 1169 LLCPTGHAV 1177

RESULT 5

S40770

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; PC1285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990

A:title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116

A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Product: nucleotide-binding motif A (P-loop)

F:1312-1317/Product: nucleotide-binding motif B

F:1316-1319/Product: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match

Best Local Similarity 92.0%; Score 46; DB 1; Length 3011;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

|||||

Db 1169 LLCPAGHAV 1177

RESULT 6

GNWVCJ

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001

C:Accession: A38465

R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991

A:title: Structure and organization of the hepatitis C virus genome isolated from human

A:Reference number: A38465; MUID:91140698

A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>

A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Product: nucleotide-binding motif A (P-loop)
F:1312-1317/Product: nucleotide-binding motif B
F:1316-1319/Product: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 88.0%; Score 44; DB 1; Length 3010;

Best Local Similarity 88.9%; Pred. No. 14;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

|||||

Db 1169 LLCPTGHAV 1177

RESULT 7

GNWVCJ

genome polyprotein - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001

C:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:title: Molecular cloning of the human hepatitis C virus genome from Japanese patient

A:Reference number: A39253; MUID:91088550

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KAT2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polyprotein have not been determined.

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Product: nucleotide-binding motif A (P-loop)

F:1312-1317/Product: nucleotide-binding motif B

F:1316-1319/Product: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224

Query Match 86.0%; Score 43; DB 1; Length 3010;

Best Local Similarity 77.8%; Pred. No. 21;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9

|||||

Db	1169	LLCPSGHV 1177	
RESULT	8		
GNMVTW			
genome polyprotein - hepatitis C virus (strain Taiwan)			
N:Contains:	capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5		
C:Species:	hepatitis C virus		
A:Note:	host Homo sapiens (man)		
C:Date:	31-Dec-1992	#sequence_revision 31-Dec-1992	#text_change 19-Jan-2001
C:Accession:	A40244		
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.			
Virology	188, 102-113, 1992		
A:Title:	The Taiwanese hepatitis C virus genome: sequence determination and mapping the		
A:Reference number:	A40244; PMID:92230206		
A:Accession:	A40244		
A:Molecule type:	genomic RNA		
A:Residues:	1-3010 <CHE>		
A:Cross-references:	GB:M84754		
C:Superfamily:	hepatitis C virus genome polyprotein		
C:Keywords:	ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural		
F:1-115/Product:	capsid protein M #status predicted <CPC>		
F:116-191/Product:	envelope protein M #status predicted <EPM>		
F:192-389/Product:	major envelope protein E #status predicted <MEE>		
F:730-1006/Product:	nonstructural protein NS1 #status predicted <NS1>		
F:1007-1615/Product:	hepatitis C virus NS2 #status predicted <NS2>		
F:1230-1237/Region:	nucleotide-binding motif A (P-loop)		
F:1312-1317/Region:	nucleotide-binding motif B		
F:1316-1319/Region:	DEXH motif		
F:1616-1862/Product:	nonstructural protein NS4a #status predicted <N4A>		
F:1863-2013/Product:	nonstructural protein NS4b #status predicted <N4B>		
F:2014-3010/Product:	nonstructural protein NS5 #status predicted <NS5>		
F:196.209.233.234.250.305.325.417.423.430.448.532.540.556.576.623.645.1213.1255.2041.207			
Query Match	86.0%;	Score 43;	DB 1; Length 3010;
Best Local Similarity	77.8%;	Pred. No. 21;	
Matches	7; Conservative	1; Mismatches	1; Indels 0; Gaps 0;
QY	1	LLCPAGHAV 9	
Db	1169	LLCPSGHV 1177	
RESULT	9		
J01303			
genome polyprotein - hepatitis C virus (isolate HC-J6)			
N:Contains:	capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5		
C:Species:	hepatitis C virus		
C:Date:	19-May-2000	#sequence_revision 19-May-2000	#text_change 17-Nov-2000
C:Accession:	J01303		
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.			
J. Gen. Virol.	72, 2697-2704, 1991		
A:Title:	Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human		
A:Reference number:	J01303; PMID:92044440		
A:Accession:	J01303		
A:Molecule type:	genomic RNA		
A:Residues:	1-3033 <OKA>		
A:Cross-references:	GB:D00944; NID:g221650; PIDN:BA00792.1; PID:g221651		
A:Experimental source:	isolate HC-J6 from a Japanese individual		
C:Superfamily:	hepatitis C virus genome polyprotein		
C:Keywords:	ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane		
F:1-115/Product:	capsid protein C #status predicted <CPC>		
F:116-191/Product:	envelope protein M #status predicted <EPM>		
F:192-389/Product:	major envelope protein E #status predicted <MEE>		
F:730-1010/Product:	nonstructural protein NS1 #status predicted <NS1>		
F:1011-1619/Product:	hepatitis C virus NS2 #status predicted <NS2>		
F:1316-1321/Region:	nucleotide-binding motif B		
F:1320-1323/Region:	DEXH motif		
F:1620-1866/Product:	nonstructural protein NS4a #status predicted <N4A>		
F:1867-2017/Product:	nonstructural protein NS4b #status predicted <N4B>		
F:2018-3033/Product:	nonstructural protein NS5 #status predicted <NS5>		
F:196.209.234.305.325.417.423.430.448.477.534.542.558.578.627.649.1091.1217.1259.2038			
Query Match	84.0%;	Score 42;	DB 1; Length 3033;
Best Local Similarity	77.8%;	Pred. No. 31;	
Matches	7; Conservative	1; Mismatches	1; Indels 0; Gaps 0;
QY	1	LLCPAGHAV 9	
Db	1173	VLCPRGHV 1181	
RESULT	10		
JC5620			
genome polyprotein - hepatitis C virus (isolate EUH1480)			
N:Contains:	capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5		
C:Species:	hepatitis C virus		
C:Date:	19-May-2000	#sequence_revision 19-May-2000	#text_change 19-Jan-2001
C:Accession:	JC5620		
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.			
Biochem. Biophys. Res. Commun.	236, 44-49, 1997		
A:Title:	The complete coding sequence of hepatitis C virus genotype 5a, the predominant		
A:Reference number:	JC5620; PMID:97366593		
A:Accession:	JC5620		
A:Molecule type:	mRNA		
A:Residues:	1-3014 <CHA>		
A:Cross-references:	GB:Y13184		
A:Experimental source:	genotype 5a, which predominates in South Africa		
A:Note:	the translation of the nucleotide sequence is not complete in this paper		
C:Superfamily:	hepatitis C virus genome polyprotein		
C:Keywords:	ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se		
F:2-115/Product:	capsid protein C #status predicted <CPC>		
F:116-191/Product:	envelope protein M #status predicted <EPM>		
F:192-389/Product:	major envelope protein E #status predicted <MEE>		
F:730-730/Product:	hypervariable #status predicted		
F:730-730/Product:	nonstructural protein NS1 #status predicted <NS1>		
F:731-1007/Product:	nonstructural protein NS2 #status predicted <NS2>		
F:1008-1616/Product:	hepatitis C virus NS3 #status predicted <NS3>		
F:1231-1238/Region:	nucleotide-binding motif A (P-loop)		
F:1313-1318/Region:	nucleotide-binding motif B		
F:1317-1320/Region:	DEXH motif		
F:1617-1863/Product:	nonstructural protein NS4a #status predicted <N4A>		
F:1864-2014/Product:	nonstructural protein NS4b #status predicted <N4B>		
F:2015-3014/Product:	nonstructural protein NS5 #status predicted <NS5>		
F:2210-2249/Region:	interferon sensitivity determining #status predicted		
Query Match	78.0%;	Score 39;	DB 1; Length 3014;
Best Local Similarity	55.6%;	Pred. No. 1e+02;	
Matches	5; Conservative	3; Mismatches	1; Indels 0; Gaps 0;
QY	1	LLCPAGHAV 9	
Db	1170	IMCPSGHV 1178	
RESULT	11		
G90806			
probable polyprotein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05			
C:Species:	Escherichia coli		
C:Date:	18-Jul-2001	#sequence_revision 18-Jul-2001	#text_change 03-Aug-2001
C:Accession:	G90806		
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.			
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcori, M.; Shinagawa, H.			
DNA Res.	8, 11-22, 2001		
A:Title:	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g		
A:Reference number:	A99629; PMID:21156231; PMID:11258796		
A:Accession:	G90806		
A:Status:	preliminary		

A:Molecule type: DNA
A:Residues: 1-177 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834846.1; PID:g13360887; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECs1423
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 76.0%; Score 38; DB 2; Length 177;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CPAGHAV 9
|||
Db 60 CPTGHAV 66

RESULT 12
C85666
probable polyprotein Z1679 [imported] - Escherichia coli (strain O157:H7, substrain EDL958)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85666
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85666
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AE005174; NID:g12514577; PIDN:AAG55791.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1679
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 76.0%; Score 38; DB 2; Length 177;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CPAGHAV 9
|||
Db 60 CPTGHAV 66

RESULT 13
B64847
probable polyprotein b1045 [similarity] - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 02-Nov-2001
C:Accession: B64847
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B64847
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Residues: 1-177 <BLAT>
A:Cross-references: GB:AE000206; GB:U00096; NID:g1787282; PIDN:AAC74129.1; PID:g1787283;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 76.0%; Score 38; DB 2; Length 177;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CPAGHAV 9
|||
Db 60 CPTGHAV 66

RESULT 14
AG2780

hypothetical protein Atul660 [imported] - Agrobacterium tumefaciens (strain C58, Dupo
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AG2780
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCL
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2780
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42661.1; PID:g17740094; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul660
A:Map position: circular chromosome

Query Match 76.0%; Score 38; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LCPAGH 7
|||
Db 85 LCPAGH 90

RESULT 15
S35092

plakoglobin - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Mar-1999
C:Accession: S35092
R:Butz, S.; Stappert, J.; Weissig, H.; Kemler, R.
Science 257, 1142-1144, 1992
A:Title: Plakoglobin and beta-catenin: distinct but closely related.
A:Reference number: S35091; MUID:92376536
A:Accession: S35092
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-621 <BUT>
A:Cross-references: EMBL:M90365
C:Keywords: cytoskeleton

Query Match 72.0%; Score 36; DB 2; Length 621;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LCPAGHA 8
|||
Db 386 LCPANHA 392

Search completed: August 23, 2002, 09:56:47
Job time: 137 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:21 ; Search time 37 seconds
(without alignments)

9,418 Million cell updates/sec

Title: US-08-854-825-26

Perfect score: 50

Sequence: 1 LLCAGHAV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	3011	1	POLG_HCV1
2	47	94.0	3010	1	POLG_HCVJT
3	46	92.0	3011	1	POLG_HCVBK
4	44	88.0	3010	1	POLG_HCVBK
5	43	86.0	3010	1	POLG_HCVJA
6	43	86.0	3010	1	POLG_HCVTW
7	42	84.0	3033	1	POLG_HCVJ6
8	37	74.0	295	1	MYRA_MICGR
9	36	72.0	621	1	PLAK_MOUSE
10	36	72.0	738	1	PLAK_XENLA
11	36	72.0	743	1	PLAK_HUMAN
12	36	72.0	781	1	CTNB_HUMAN
13	36	72.0	781	1	CTNB_MOUSE
14	36	72.0	781	1	CTNB_RAT
15	36	72.0	781	1	CTNB_XENLA
16	36	72.0	818	1	CTNB_URECA
17	36	72.0	843	1	ARM_DROME
18	36	72.0	972	1	UVRA_MYCTU
19	36	72.0	988	1	PCK1_SCHPO
20	35	70.0	331	1	YHBU_ECOLI
21	35	70.0	2499	1	MPRI_BOVIN
22	34	68.0	417	1	TR12_HUMAN
23	34	68.0	440	1	T10B_HUMAN
24	34	68.0	507	1	Y145_CAEEL
25	34	68.0	525	1	VCL_THICC
26	34	68.0	529	1	PGL2_RALSO
27	34	68.0	531	1	PGL1_RALSO
28	34	68.0	540	1	BAIC_EUBSP
29	34	68.0	1376	1	CRBH_HUMAN
30	34	68.0	2871	1	FBNI_BOVIN
31	34	68.0	2871	1	FBNI_HUMAN
32	34	68.0	2871	1	FBNI_MOUSE
33	34	68.0	2871	1	FBNI_PIG

34 34 68.0 2907 1 FBN2_MOUSE
35 34 68.0 2911 1 FBN2_HUMAN
36 34 68.0 3033 1 POLG_HCVJ8
37 33 66.0 152 1 YCHJ_ECOLI
38 33 66.0 152 1 YCHJ_SHIFL
39 33 66.0 219 1 YJL8_YEAST
40 33 66.0 332 1 ISPA_RHISN
41 33 66.0 379 1 FLIP_BUCAI
42 33 66.0 419 1 SEP2_DROME
43 33 66.0 677 1 SP87_DICDI
44 33 66.0 764 1 YJJO_YEAST
45 33 66.0 1174 1 RPB2_HUMAN

ALIGNMENTS

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OC Hepacivirus
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M62321; AAA45676.1; -;
CC PIR: A39166; GNMVC3.
CC HSP: P27958; IHEI.
CC MEROPS: S29.001; -;
CC MEROPS: U39.001; -;
CC InterPro: IPR001410; DEAD
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.

```

DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RDRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF0186062; HCV_NS1; 1.
DR SMART; SM00492; HelicC3; 1.
DR Polyprotein: Glycoprotein; Transfrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 50; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPPAGHAV 9
DB 1169 LLCPPAGHAV 1177

RESULT 2

```

```

POLG_HCVJT STANDARD; PRT; 3010 AA.
ID POLG_HCVJT
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -! FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -! CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -! SUBUNIT: THE ENVELOPE OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -! SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11168; BAA01943.1;
CC PIR: A45573; A45573.
CC HSP: P26663; 1JXP.
CC MEROPS: S29.001; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RDRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; HCV_RDRP; 1.

```

DR Pfam; PF00271; helicase_C; 1.
 DR SMART; PD186062; HCV_NS1; 1.
 DR SMART; SM00492; HELIC3; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 94.08; Score 47; DB 1; Length 3010;
 Best Local Similarity 88.98; Pred. No. 1.6;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLCPAGHAV 9
 |||||
 Db 1169 LLCPSGHAV 1177

RESULT 3
 ID POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;

[1] SEQUENCE FROM N.A.
 RP MEDLINE-92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE-97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98154321; PubMed=9493270;
 RA Kim J.-L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1',
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1; -.
 DR PIR; A36814; GNMVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1AIV; 16-FEB-99.
 DR MEROPS; S29.001; -.
 DR -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Prodom: PD186062; HCV_NS1; 1.
DR SMART: SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 747 809 PROTEIN P7.
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT CHAIN 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 92.0%; Score 46; DB 1; Length 3011;
Best Local Similarity 88.9%; Pred. No. 2.4;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLCPAGHAV 9
Db 1169 LLCPTGHAV 1177

RESULT 4
POLG_HCVBK
ID POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.:
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetcky L.:
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parage H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomska Z.:
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.:
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58335; AAA72945.1; -.
DR PIR: A38465; GNVVTC.
DR PDB: 1AIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRp; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327189 MW; P8422D5ECCFDFD9C CRC64;

Query Match 88.0%; Score 44; DB 1; Length 3010;
Best Local Similarity 88.9%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLCPCGHAV 9
Db 1169 LLCPCGHAV 1177

RESULT 5
POLG_HCVJA
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genomé polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RN DISCUSSION OF SEQUENCE.
RP MEDLINE=91192160; PubMed=1849488;
RX Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in Pl and Ser or Ala in Pl.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90208; BAA14233.1; -.
DR PIR; A39253; GNVVCJ.
DR HSP; P26663; LXP.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRp; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.

DR SMART: SM00492; HELIC3; 1.
 KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 136 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 86.0%; Score 43; DB 1; Length 3010;
 Best Local Similarity 77.8%; Pred. No. 8.1;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCAGHAV 9
 DB 1169 LLCPSGHV 1177

RESULT 6
 POLG_HCVTW
 ID POLG_HCVTW STANDARD: PRT: 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2;
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RL mapping the 5' termini of viral genomic and antigenomic RNA.";
 CC Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M84754; -; NOT_ANNOTATED_CDS.
 CC PIR: A40244; GNVVTV.
 CC HSSP: P26663; LJXP.
 CC MEROPS: S29.001; -.
 CC MEROPS: U39.001; -.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR00745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RGRP.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; HCV_RGRP; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 136 196

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 86.0%; Score 43; DB 1; Length 3010;
 Best Local Similarity 77.8%; Pred. No. 8.1;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPSGHV 9
 Db 1169 LLCPSGHV 1177

RESULT 7
 POLG_HCVJ6 STANDARD; PRT; 3033 AA.

AC P26660; 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contents: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9204440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00944; BAA00792.1; -.
 DR PIR; JQ1303; JQ1303.
 DR HSSP; P27958; IHEI.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RdRp; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 1 115 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 734 1010 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1234 1241 ATP (POTENTIAL).
 FT SITE 1320 1323 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match      84.0%; Score 42; DB 1; Length 3033;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPCAGHAV 9
   :||| ||||
Db 1173 VLCPCGHAV 1181

RESULT 8
MYRA_MICGR
AC P37000; STANDARD; PRT; 295 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Mycinamicin-resistance protein myra.
GN MYRA.
OS Micromonospora griseorubida.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micromonosporineae; Micromonosporaceae;
OC Micromonospora.
OX NCBI_TaxID=28040;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94215886; PubMed=8163173;
RA Inouye M., Morohoshi T., Horinouchi S., Beppu T.;
RT "Cloning and sequences of two macrolide-resistance-encoding genes
   from mycinamicin-producing Micromonospora griseorubida.";
RL Gene 141:39-46(1994).
CC -!- FUNCTION: CONFERS STRONG RESISTANCE TO MYCINAMICIN (MM) AND
   TYLOSIN (TY).
CC -!- SIMILARITY: TO E.COLI RMA AND B.SUBTILIS YXJB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D16099; BAA03674.1; -
KW Antibiotic resistance.
SQ SEQUENCE 295 AA; 31348 MW; AC42FOAF750BCB1A CRC64;

Query Match      74.0%; Score 37; DB 1; Length 295;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLCPCAGHA 8
   | |||||
Db 29 LRPCAGHS 36

RESULT 9
PLAK_MOUSE
AC Q02257; STANDARD; PRT; 621 AA.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Junction plakoglobin (Desmoplakin III) (fragment).
GN JUP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=92376536; PubMed=1509266;
RA Butz S., Stappert J., Weissig H., Kemler R.;
RT "Plakoglobin and beta-catenin: distinct but closely related.";
RL Science 257:1142-1144(1992).
RN [2]
RP REVISIONS TO 294 AND 296.
RA Butz S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMMON JUNCTIONAL PLAQUE PROTEIN. THE MEMBRANE-
   ASSOCIATED PLAQUES ARE ARCHITECTURAL ELEMENTS IN AN IMPORTANT
   STRATEGIC POSITION TO INFLUENCE THE ARRANGEMENT AND FUNCTION
   OF BOTH THE CYTOSKELETON AND THE CELLS WITHIN THE TISSUE. THE
   PRESENCE OF PLAKOGLOBIN IN BOTH THE DESMOSOMES AND IN THE
   INTERMEDIATE JUNCTIONS SUGGESTS THAT IT PLAYS A CENTRAL ROLE IN
   THE STRUCTURE AND FUNCTION OF SUBMEMBRANOUS PLAQUES.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC IN A SOLUBLE AND MEMBRANE-
   ASSOCIATED FORM.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS AT LEAST 9 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90365; AAB02885.1; -
DR PIR; S35092; S35092.
DR HSP; Q02248; IDOW.
DR MGD; MGI:96650; Jup.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 11.
DR SMART; SM00185; ARM; 8.
DR PROSITE; PS0176; ARM_REPEAT; 9.
KW Cell adhesion; Cytoskeleton; Structural protein; Repeat.
FT NON_TER 1 47 ARM 1.
FT REPEAT 8 47 ARM 2.
FT REPEAT 92 131 ARM 3.
FT REPEAT 134 173 ARM 4.
FT REPEAT 218 257 ARM 5.
FT REPEAT 259 296 ARM 6.
FT REPEAT 299 340 ARM 7.
FT REPEAT 346 386 ARM 8.
FT REPEAT 388 427 ARM 9.
FT REPEAT 450 489 ARM 9.
SQ SEQUENCE 621 AA; 68111 MW; 17CF444607422BAA CRC64;

Query Match      72.0%; Score 36; DB 1; Length 621;
Best Local Similarity 85.7%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LCPAGHA 8
   |||||
Db 386 LCPANHA 392

RESULT 10
PLAK_XENLA
ID PLAK_XENLA STANDARD; PRT; 738 AA.
AC P30998;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Junction plakoglobin (Desmoplakin III).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.

```

OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93093332; PubMed=1459359;
RA Fouquet B., Zimbelmann R., Franke W.W.;
RT "Identification of plakoglobin in oocytes and early embryos of
RT Xenopus laevis: maternal expression of a gene encoding a junctional
RT plaque protein.";
RL Differentiation 51:187-194(1992).
RN [2]
RP SEQUENCE OF 133-292 FROM N.A.
RX MEDLINE=93012479; PubMed=1397690;
RA de Marais A.A., Moon R.T.;
RT "The armadillo homologs beta-catenin and plakoglobin are
RT differentially expressed during early development of Xenopus
RT laevis.";
RL Dev. Biol. 153:337-346(1992).
CC -!- FUNCTION: COMMON JUNCTIONAL PLAQUE PROTEIN. THE MEMBRANE-
CC ASSOCIATED PLAQUES ARE ARCHITECTURAL ELEMENTS IN AN IMPORTANT
CC STRATEGIC POSITION TO INFLUENCE THE ARRANGEMENT AND FUNCTION
CC OF BOTH THE CYTOSKELETON AND THE CELLS WITHIN THE TISSUE. THE
CC PRESENCE OF PLAKOGLOBIN IN BOTH THE DESMOSOMES AND IN THE
CC INTERMEDIATE JUNCTIONS SUGGESTS THAT IT PLAYS A CENTRAL ROLE IN
CC THE STRUCTURE AND FUNCTION OF SUBMEMBRANOUS PLAQUES.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC IN A SOLUBLE AND MEMBRANE-
CC ASSOCIATED FORM.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95593; AAA49931.1; -;
DR EMBL; X67078; CAA47463.1; -;
DR PIR; S24636; S24636.
DR PIR; S35093; S35093.
DR HSPP; Q02248; 2BCT.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 11.
DR SMART; SM00185; Arm; 10.
DR PROSITE; PS0176; ARM_REPEAT; 9.
KW Cell adhesion; Cytoskeleton; Structural protein; Repeat.
FT REPEAT 128 167 ARM 1.
FT REPEAT 212 251 ARM 2.
FT REPEAT 254 293 ARM 3.
FT REPEAT 338 377 ARM 4.
FT REPEAT 379 416 ARM 5.
FT REPEAT 419 460 ARM 6.
FT REPEAT 466 506 ARM 7.
FT REPEAT 508 547 ARM 8.
FT REPEAT 570 609 ARM 9.
FT REPEAT 611 650 ARM 10.
FT REPEAT 185 185 I -> V (IN REF. 2).
FT CONFLICT 226 226 A -> T (IN REF. 2).
SQ SEQUENCE 738 AA; 81711 MW; 569DBE69D08B8C58 CRC64;

Query Match 72.08; Score 36; DB 1; Length 738;
Best Local Similarity 85.78; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LCPAGHA 8
Db 506 LCPANHA 512
|||||

RESULT 11

PLAK_HUMAN
ID PLAK_HUMAN STANDARD; PRT; 743 AA.
AC P14923;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Junction plakoglobin (Desmoplakin III).
GN JUP OR DP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89264555; PubMed=2726765;
RA Franke W.W., Goldschmidt M.D., Zimbelmann R., Mueller H.M.,
RA Schiller D.L., Cowlin P.;
RT "Molecular cloning and amino acid sequence of human plakoglobin, the
RT common junctional plaque protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4027-4031(1989).
CC -!- FUNCTION: COMMON JUNCTIONAL PLAQUE PROTEIN. THE MEMBRANE-
CC ASSOCIATED PLAQUES ARE ARCHITECTURAL ELEMENTS IN AN IMPORTANT
CC STRATEGIC POSITION TO INFLUENCE THE ARRANGEMENT AND FUNCTION
CC OF BOTH THE CYTOSKELETON AND THE CELLS WITHIN THE TISSUE. THE
CC PRESENCE OF PLAKOGLOBIN IN BOTH THE DESMOSOMES AND IN THE
CC INTERMEDIATE JUNCTIONS SUGGESTS THAT IT PLAYS A CENTRAL ROLE IN
CC THE STRUCTURE AND FUNCTION OF SUBMEMBRANOUS PLAQUES.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC IN A SOLUBLE AND MEMBRANE-
CC ASSOCIATED FORM.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 8 ARM REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23410; AAA64895.1; -;
DR PIR; A32905; A32905.
DR HSPP; Q02248; 2BCT.
DR MIM; 173325; -;
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 10.
DR SMART; SM00185; ARM; 8.
DR PROSITE; PS0176; ARM_REPEAT; 8.
KW Cell adhesion; Cytoskeleton; Structural protein; Repeat.
FT INIT_MET 0 0 PROBABLE.
FT REPEAT 141 179 ARM 1.
FT REPEAT 183 224 ARM 2.
FT REPEAT 225 252 ARM 3.
FT REPEAT 308 351 ARM 4.
FT REPEAT 389 430 ARM 5.
FT REPEAT 431 473 ARM 6.
FT REPEAT 478 521 ARM 7.
FT REPEAT 582 624 ARM 8.
FT CONFLICT 96 99 GOL A -> DSSL (IN REF. 1; AAA64895).
FT CONFLICT 139 139 V -> A (IN REF. 1; AAA64895).
SQ SEQUENCE 743 AA; 81498 MW; 472741F400D388FD CRC64;

Query Match 72.08; Score 36; DB 1; Length 743;
Best Local Similarity 85.78; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LCPAGHA 8
Db 508 LCPANHA 514
|||||

```

RESULT 12
CTNB_HUMAN
ID CTNB_HUMAN STANDARD; PRT; 781 AA.
AC P35222;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-catenin.
GN CTNNB1 OR CTNNB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP MEDLINE-20145417; PubMed-10679188;
RA Kikuchi A.;
RT Regulation of beta-catenin signaling in the Wnt pathway.*;
RL Biochem. Biophys. Res. Commun. 268:243-248(2000).
[3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 133-664.
RX MEDLINE-20578806; PubMed-11136974;
RA Graham T.A., Weaver C., Mao F., Kimelman D., Xu W.;
RT "Crystal structure of a beta-catenin/Tcf complex.*;
RL Cell 103:885-896(2000).
CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CELL ADHESION AND IN
CC SIGNAL TRANSDUCTION THROUGH THE WNT PATHWAY.
CC -!- SUBUNIT: THE CYTOSOLIC FORM BOUND TO AXIN ALLOWS THE STIMULATION
CC OF ITS DEGRADATION. THE CYTOSOLIC FORM ALSO INTERACTS WITH APC AND
CC E-CADHERIN. THE NUCLEAR FORM BINDS TCF/LEF-1 AND MAY ALSO BIND
CC PONTIN32 AND DUPLIN. CAN HETERODIMERIZE WITH ALPHA-CATENIN AND
CC ALSO FORM A TRIMERIC COMPLEX WITH CADHERIN. MEDIATES THE ANCHORAGE
CC OF THE CADHERIN COMPLEX TO THE CORTICAL ACTIN CYTOSKELETON.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC WHEN IT IS UNSTABILIZED (HIGH
CC LEVEL OF PHOSPHORYLATION). TRANSLOCATES TO THE NUCLEUS WHEN IT IS
CC STABILIZED (LOW LEVEL OF PHOSPHORYLATION).
CC -!- PTM: PROBABLY PHOSPHORYLATED BY GSK-3B. WNT-DEPENDENT ACTIVATION
CC OF DVL ANTAGONIZES THE ACTION OF GSK-3B. PHOSPHORYLATION OF BETA-
CC CATENIN IS REDUCED AND THE PROTEIN IS NO LONGER DEGRADED,
CC RESULTING IN ITS ACCUMULATION IN CYTOPLASM.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 12 ARM REPEATS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: X67838; CAA61107.1; -
DR EMBL: 219054; CAA79497.1; -
DR PIR: S31988; S31988.
DR PDB: 1G3J; 17-JAN-01.
DR TRANSFAC: T02872; -.
DR MIM: 116806; -.
DR InterPro: IPR000225; Armadillo.
DR Pfam: PF00514; Armadillo_seg; 12.
DR SMART: SM00185; ARM; 11.
DR PROSITE: PS50176; ARM_REPEAT; 9.
KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW Repeat; 3D-structure.
FT REPEAT 151 191 ARM 1.
FT REPEAT 193 234 ARM 2.

```

```

FT REPEAT 235 276 ARM 3.
FT REPEAT 277 318 ARM 4.
FT REPEAT 319 360 ARM 5.
FT REPEAT 361 389 ARM 6.
FT REPEAT 400 441 ARM 7.
FT REPEAT 442 484 ARM 8.
FT REPEAT 489 530 ARM 9.
FT REPEAT 531 571 ARM 10.
FT REPEAT 594 636 ARM 11.
FT REPEAT 637 666 ARM 12.
SQ SEQUENCE 781 AA; 85496 MW; CB78F165A3EEF86E CRC64;

Query Match 72.0%; Score 36; DB 1; Length 781;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LCPAGHA 8
DB 519 LCPANHA 525
|||||

RESULT 13
CTNB_MOUSE
ID CTNB_MOUSE STANDARD; PRT; 781 AA.
AC Q02248;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-catenin.
GN CTNNB1 OR CATNB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP MEDLINE-92376536; PubMed-1509266;
RA Butz S., Stappert J., Weissig H., Kemler R.;
RT "Plakoglobin and beta-catenin: distinct but closely related.*;
RL Science 257:1142-1144(1992).
[2]
RP REVIEW.
RX MEDLINE-20145417; PubMed-10679188;
RA Kikuchi A.;
RT "Regulation of beta-catenin signaling in the Wnt pathway.*;
RL Biochem. Biophys. Res. Commun. 268:243-248(2000).
[3]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 150-665.
RX MEDLINE-97442350; PubMed-9298899;
RA Huber A.H., Nelson W.J., Weis W.I.;
RT "Three-dimensional structure of the armadillo repeat region of beta-
RL catenin.*;
RL Cell 90:871-882(1997).
[4]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 118-149 IN COMPLEX WITH
RP CTNNB1.
RX MEDLINE-20337986; PubMed-10882138;
RA Pokutta S., Weis W.I.;
RT "Structure of the dimerization and beta-catenin-binding region of
RL alpha-catenin.*;
RL Mol. Cell 5:533-543(2000).
CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CELL ADHESION AND IN
CC SIGNAL TRANSDUCTION THROUGH THE WNT PATHWAY.
CC -!- SUBUNIT: THE CYTOSOLIC FORM BOUND TO AXIN ALLOWS THE STIMULATION
CC OF ITS DEGRADATION. THE CYTOSOLIC FORM ALSO INTERACTS WITH APC AND
CC E-CADHERIN. THE NUCLEAR FORM BINDS TCF/LEF-1 AND MAY ALSO BIND
CC PONTIN32 AND DUPLIN. CAN HETERODIMERIZE WITH ALPHA-CATENIN AND
CC ALSO FORM A TRIMERIC COMPLEX WITH CADHERIN. MEDIATES THE ANCHORAGE
CC OF THE CADHERIN COMPLEX TO THE CORTICAL ACTIN CYTOSKELETON.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC WHEN IT IS UNSTABILIZED (HIGH
CC LEVEL OF PHOSPHORYLATION). TRANSLOCATES TO THE NUCLEUS WHEN IT IS
CC STABILIZED (LOW LEVEL OF PHOSPHORYLATION).

```

```
CC -!- PTM: PROBABLY PHOSPHORYLATED BY GSK-3B. WNT-DEPENDENT ACTIVATION
CC OF DVL ANTAGONIZES THE ACTION OF GSK-3B. PHOSPHORYLATION OF BETA-
CC CATEININ IS REDUCED AND THE PROTEIN IS NO LONGER DEGRADED,
CC RESULTING IN ITS ACCUMULATION IN CYTOPLASM.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 12 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M50364; AAA37280.1; -.
CC PIR: S35091; S35091.
CC PDB: 2BCT; 15-OCT-97.
CC PDB: 3BCT; 19-NOV-97.
CC PDB: 1DOW; 12-JUL-00.
CC TRANSFAC: T02984; -.
CC MGD: MGI:88276; Catnb.
CC InterPro: IPR000225; Armadillo.
CC Pfam: PF00514; Armadillo_seg; 12.
CC SMART: SM00185; ARM; 11.
CC PROSITE: PS50176; ARM_REPEAT; 9.
CC Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW Repeat; 3D-structure.
FT REPEAT 151 191 ARM 1.
FT REPEAT 193 234 ARM 2.
FT REPEAT 235 276 ARM 3.
FT REPEAT 277 318 ARM 4.
FT REPEAT 319 360 ARM 5.
FT REPEAT 361 389 ARM 6.
FT REPEAT 400 441 ARM 7.
FT REPEAT 442 484 ARM 8.
FT REPEAT 489 530 ARM 9.
FT REPEAT 531 571 ARM 10.
FT REPEAT 594 636 ARM 11.
FT REPEAT 637 666 ARM 12.
FT REPEAT 781 AA; 85470 MW; D708F170A3FBED6E CRC64;
SQ SEQUENCE 781 AA; 85470 MW; D708F170A3FBED6E CRC64;

Query Match 72.0%; Score 36; DB 1; Length 781;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LCPAGHA 8
DB 519 LCPANHA 525
|||||

RESULT 14
CTNB_RAT CTNB_RAT STANDARD; PRT; 781 AA.
ID CTNB_RAT
AC Q9WU82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-catenin.
GN CTNNB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RA MEDLINE=99428593; PubMed=10497305;
RA Chung S.S.W., Lee W.M., Cheng C.Y.;
RT "Study on the formation of specialized inter-Sertoli cell junctions in
RT vitro.";
RL J. Cell. Physiol. 181:258-272(1999).
```

```
CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CELL ADHESION AND IN
CC SIGNAL TRANSDUCTION THROUGH THE WNT PATHWAY (BY SIMILARITY).
CC SUBUNIT: THE CYTOSOLIC FORM BOUND TO AXIN ALLOWS THE STIMULATION
CC OF ITS DEGRADATION. THE CYTOSOLIC FORM ALSO INTERACTS WITH APC AND
CC E-CADHERIN. THE NUCLEAR FORM BINDS TCF/LEF-1 AND MAY ALSO BIND
CC PONTIN2 AND DUPLIN. CAN HETERODIMERIZE WITH ALPHA-CATENIN AND
CC ALSO FORM A TRIMERIC COMPLEX WITH CADHERIN. MEDIATES THE ANCHORAGE
CC OF THE CADHERIN COMPLEX TO THE CORTICAL ACTIN CYTOSKELETON (BY
CC SIMILARITY).
CC SUBCELLULAR LOCATION: CYTOPLASMIC WHEN IT IS UNSTABILIZED (HIGH
CC LEVEL OF PHOSPHORYLATION). TRANSLOCATES TO THE NUCLEUS WHEN IT IS
CC STABILIZED (LOW LEVEL OF PHOSPHORYLATION) (BY SIMILARITY).
CC TISSUE SPECIFICITY: EXPRESSED IN THE TESTIS.
CC -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED AT E30-60 DAY DPC IN THE
CC TESTIS. REDUCED EXPRESSION AT E90 DAY DPC.
CC -!- PTM: PROBABLY PHOSPHORYLATED BY GSK-3B. WNT-DEPENDENT ACTIVATION
CC OF DVL ANTAGONIZES THE ACTION OF GSK-3B. PHOSPHORYLATION OF BETA-
CC CATEININ IS REDUCED AND THE PROTEIN IS NO LONGER DEGRADED,
CC RESULTING IN ITS ACCUMULATION IN CYTOPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 12 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF121265; AAD28504.1; -.
CC HSSP: Q02248; 2BCT.
CC InterPro: IPR000225; Armadillo.
CC Pfam: PF00514; Armadillo_seg; 12.
CC SMART: SM00185; ARM; 11.
CC PROSITE: PS50176; ARM_REPEAT; 9.
CC Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW Repeat.
FT REPEAT 151 191 ARM 1.
FT REPEAT 193 234 ARM 2.
FT REPEAT 235 276 ARM 3.
FT REPEAT 277 318 ARM 4.
FT REPEAT 319 360 ARM 5.
FT REPEAT 361 389 ARM 6.
FT REPEAT 400 441 ARM 7.
FT REPEAT 442 484 ARM 8.
FT REPEAT 489 530 ARM 9.
FT REPEAT 531 571 ARM 10.
FT REPEAT 594 636 ARM 11.
FT REPEAT 637 666 ARM 12.
FT REPEAT 781 AA; 85454 MW; 9C29186B6DD54B87 CRC64;
SQ SEQUENCE 781 AA; 85454 MW; 9C29186B6DD54B87 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 781;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LCPAGHA 8
DB 519 LCPANHA 525
|||||

RESULT 15
CTNB_XENLA CTNB_XENLA STANDARD; PRT; 781 AA.
ID CTNB_XENLA
AC P26233;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-catenin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae.
```

```

OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=92073903; PubMed=1962194;
RA McCrea P.D., Turk C.W., Gumbiner B.M.;
RT "A homolog of the armadillo protein in Drosophila (plakoglobin)
RL associated with E-cadherin.";
RL Science 254:1359-1361(1991).
CC -!- FUNCTION: BINDS TO THE CYTOPLASMIC DOMAIN OF THE CELL-CELL
CC ADHESION MOLECULE E-CADHERIN, AND PERHAPS TO OTHER (MEMBRANE)
CC PROTEINS. THE ASSOCIATION OF CATENINS TO CADHERINS PRODUCES A
CC COMPLEX WHICH IS LINKED TO THE ACTIN FILAMENT NETWORK, AND WHICH
CC SEEMS TO BE OF PRIMARY IMPORTANCE FOR CADHERINS CELL-ADHESION
CC PROPERTIES. ESSENTIAL COMPONENT OF THE VERTEBRATE WNT-SIGNALING
CC PATHWAY AND IS REQUIRED FOR FORMATION OF THE EMBRYONIC AXIS.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77013; AAA49670.1; -.
DR PIR: S35099; S35099.
DR HSP: Q02248; 2BCT.
DR TRANSFAC: T03026; -.
DR InterPro: IPR000225; Armadillo.
DR Pfam: PF00514; Armadillo_seg; 12.
DR SMART: SM00185; ARM: 11.
DR PROSITE: PS50176; ARM_REPEAT; 9.
KW Cell adhesion; Cytoskeleton; Structural protein; Repeat.
FT REPEAT 141 180 ARM 1.
FT REPEAT 225 264 ARM 2.
FT REPEAT 267 306 ARM 3.
FT REPEAT 351 390 ARM 4.
FT REPEAT 391 429 ARM 5.
FT REPEAT 432 473 ARM 6.
FT REPEAT 479 519 ARM 7.
FT REPEAT 521 562 ARM 8.
FT REPEAT 584 623 ARM 9.
FT REPEAT 625 664 ARM 10.
SQ SEQUENCE 781 AA; 85449 MW; 3ECD2723239F799 CRC64;

```

```

Query Match 72.0%; Score 36; DB 1; Length 781;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LCPAGHA 8
Db 519 LCPANHA 525

```

Search completed: August 23, 2002, 09:55:22
Job time: 52 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:21:58 ; Search time 139.83 Seconds
(without alignments)
11.135 Million cell updates/sec

Title: US-08-854-825-26
Perfect score: 50
Sequence: 1 LLCPCAGHAV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	50	100.0	181	12 Q91RT9	Q91rt9 hepatitis c
2	50	100.0	181	12 Q91RT8	Q91rt8 hepatitis c
3	50	100.0	181	12 Q91RT7	Q91rt7 hepatitis c
4	50	100.0	181	12 Q91RT6	Q91rt6 hepatitis c
5	50	100.0	181	12 Q91RT5	Q91rt5 hepatitis c
6	50	100.0	181	12 Q91RT4	Q91rt4 hepatitis c
7	50	100.0	181	12 Q91RT3	Q91rt3 hepatitis c
8	50	100.0	181	12 Q91RT2	Q91rt2 hepatitis c
9	50	100.0	181	12 Q91RT1	Q91rt1 hepatitis c
10	50	100.0	181	12 Q91RT0	Q91rt0 hepatitis c
11	50	100.0	181	12 Q91RS9	Q91rs9 hepatitis c
12	50	100.0	181	12 Q91RS8	Q91rs8 hepatitis c
13	50	100.0	181	12 Q91RS7	Q91rs7 hepatitis c
14	50	100.0	181	12 Q91RS6	Q91rs6 hepatitis c
15	50	100.0	181	12 Q91RS5	Q91rs5 hepatitis c
16	50	100.0	181	12 Q91RS4	Q91rs4 hepatitis c

17	50	100.0	181	12 Q91RS3	Q91rs3 hepatitis c
18	50	100.0	181	12 Q91RS2	Q91rs2 hepatitis c
19	50	100.0	181	12 Q91RS1	Q91rs1 hepatitis c
20	50	100.0	181	12 Q91RS0	Q91rs0 hepatitis c
21	50	100.0	181	12 Q91RR9	Q91rr9 hepatitis c
22	50	100.0	181	12 Q91RR8	Q91rr8 hepatitis c
23	50	100.0	181	12 Q91RR7	Q91rr7 hepatitis c
24	50	100.0	181	12 Q91RR6	Q91rr6 hepatitis c
25	50	100.0	181	12 Q91RR5	Q91rr5 hepatitis c
26	50	100.0	181	12 Q91RR3	Q91rr3 hepatitis c
27	50	100.0	181	12 Q91RR2	Q91rr2 hepatitis c
28	50	100.0	181	12 Q91RR1	Q91rr1 hepatitis c
29	50	100.0	181	12 Q91RR0	Q91rr0 hepatitis c
30	50	100.0	181	12 Q91RQ9	Q91rq9 hepatitis c
31	50	100.0	181	12 Q91RQ8	Q91rq8 hepatitis c
32	50	100.0	2436	12 081756	081756 hepatitis c
33	50	100.0	3010	12 Q9Q1Z0	Q9q1z0 hepatitis c
34	50	100.0	3010	12 Q9Q1Y9	Q9q1y9 hepatitis c
35	50	100.0	3010	12 Q9Q3H8	Q9q3h8 hepatitis c
36	50	100.0	3011	12 Q36579	Q36579 hepatitis c
37	50	100.0	3011	12 Q36608	Q36608 hepatitis c
38	50	100.0	3011	12 Q36609	Q36609 hepatitis c
39	50	100.0	3011	12 Q36610	Q36610 hepatitis c
40	50	100.0	3011	12 Q91FE5	Q91fe5 hepatitis c
41	50	100.0	3011	12 Q9ELS8	Q9els8 hepatitis c
42	50	100.0	3011	12 Q9DIT6	Q9dit6 hepatitis c
43	50	100.0	3015	12 Q9PWX5	Q9pwx5 hepatitis c
44	50	100.0	3015	12 Q9PWU9	Q9pwu9 hepatitis c
45	50	100.0	4040	12 Q91FH8	Q91fh8 mucosal dis

ALIGNMENTS

RESULT 1

Q91RT9	PRELIMINARY;	PRT;	181 AA.
ID Q91RT9	PRELIMINARY;	PRT;	181 AA.
AC Q91RT9;			
DT 01-DEC-2001 (TrEMBLrel. 19, Created)			
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE NS3 PROTEASE (FRAGMENT).			
OS Hepatitis C virus.			
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC Hepacivirus.			
OX NCBI_TaxID=11103;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=PT.1C;			
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;			
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from Clinical Strains of the Hepatitis C Virus.";			
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF369214; AAK54539.1; -			
KW Protease.			
FT NON_TER	1		
FT NON_TER	181		
SQ SEQUENCE	181 AA;	19114 MW;	BE1D0B542F014E86 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPCAGHAV 9
|||||
Db 143 LLCPCAGHAV 151

RESULT 2

Q91RT8	PRELIMINARY;	PRT;	181 AA.
ID Q91RT8	PRELIMINARY;	PRT;	181 AA.
AC Q91RT8;			

DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PT.1H;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369215; AAK54540.1; -
KW Protease.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19115 MW; 746BF671E7BCBD2E CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
DB 143 LLCPAGHAV 151
|||||
RESULT 3
Q91RT7 ID Q91RT7 PRELIMINARY; PRT; 181 AA.
AC Q91RT7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PT.1H;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369216; AAK54541.1; -
KW Protease.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19107 MW; 1372F1F783A720CD CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
DB 143 LLCPAGHAV 151
|||||
RESULT 4
Q91RT6 ID Q91RT6 PRELIMINARY; PRT; 181 AA.
AC Q91RT6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PT.252;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369217; AAK54542.1; -
KW Protease.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19102 MW; 373DE6563C3C5959 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
DB 143 LLCPAGHAV 151
|||||
RESULT 5
Q91RT5 ID Q91RT5 PRELIMINARY; PRT; 181 AA.
AC Q91RT5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PT.4;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369218; AAK54543.1; -
KW Protease.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
DB 143 LLCPAGHAV 151
|||||
RESULT 6
Q91RT4 ID Q91RT4 PRELIMINARY; PRT; 181 AA.
AC Q91RT4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.23;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369219; AAK54544.1; -
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19059 MW; 1E53C47AE8B7E5C9 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 143 LLCPAGHAV 151
|||||

RESULT 7
Q91RT3 PRELIMINARY; PRT; 181 AA.
AC Q91RT3;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.11;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369220; AAK54545.1; -
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19116 MW; 9648807F49EB1D43 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 143 LLCPAGHAV 151
|||||

RESULT 8
Q91RT2 PRELIMINARY; PRT; 181 AA.
AC Q91RT2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=PT.12;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369221; AAK54546.1; -
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19106 MW; 354881366F3070F5 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 143 LLCPAGHAV 151
|||||

RESULT 9
Q91RT1 PRELIMINARY; PRT; 181 AA.
AC Q91RT1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.161;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369222; AAK54547.1; -
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPAGHAV 9
Db 143 LLCPAGHAV 151
|||||

RESULT 10
Q91RT0 PRELIMINARY; PRT; 181 AA.
AC Q91RT0;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.170;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
```

RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
KW EMBL; AF369223; AAK54548.1; -;
FT NON_TER 1 181
SQ SEQUENCE 181 AA; 19116 MW; 9648807F49EBID43 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPCAGHAV 9
Db 143 LLCPCAGHAV 151
|||||

RESULT 11
Q91RS9 PRELIMINARY; PRT; 181 AA.
AC Q91RS9;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.174;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369224; AAK54549.1; -;
KW Protease.
FT NON_TER 1 181
SQ SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPCAGHAV 9
Db 143 LLCPCAGHAV 151
|||||

RESULT 12
Q91RS8 PRELIMINARY; PRT; 181 AA.
AC Q91RS8;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.176;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF369225; AAK54550.1; -;
KW Protease.
FT NON_TER 1 181
SQ SEQUENCE 181 AA; 19114 MW; 574AC47AE8AE5D2 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPCAGHAV 9
Db 143 LLCPCAGHAV 151
|||||

RESULT 13
Q91RS7 PRELIMINARY; PRT; 181 AA.
AC Q91RS7;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.177;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369226; AAK54551.1; -;
KW Protease.
FT NON_TER 1 181
SQ SEQUENCE 181 AA; 19116 MW; 9648807F49EBID43 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LLCPCAGHAV 9
Db 143 LLCPCAGHAV 151
|||||

RESULT 14
Q91RS6 PRELIMINARY; PRT; 181 AA.
AC Q91RS6;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PT.183;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369227; AAK54552.1; -;
KW Protease.
FT NON_TER 1 1

FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19097 MW; 7744468EB7EFDB2E CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPCAGHAV 9
| | | | | | | | | |
Db 143 LLCPCAGHAV 151

RESULT 15
Q91RS5
ID Q91RS5 PRELIMINARY; PRT; 181 AA.
AC Q91RS5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NS3 PROTEASE (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PT.186;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369228; AAK54553.1; -.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19086 MW; ABBBD98B3ABA4D25 CRC64;

Query Match 100.0%; Score 50; DB 12; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLCPCAGHAV 9
| | | | | | | | | |
Db 143 LLCPCAGHAV 151

Search completed: August 23, 2002, 10:21:59
Job time: 1499 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:00:59 ; Search time 179.72 seconds
(without alignments)
6.180 Million cell updates/sec

Title: US-08-854-825-28
Perfect score: 45
Sequence: 1 KLUVALGINAV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	45	100.0	10	AA770080	JY-EBV cells pre-i
2	45	100.0	10	AA770080	Cytotoxic T-cell e
3	45	100.0	10	AA770080	Hepatitis C virus
4	45	100.0	10	AA770080	HCV NS3 peptide (
5	45	100.0	10	AA770080	Peptide from HCV-1
6	45	100.0	10	AA770080	T cell epitope/MHC
7	45	100.0	10	AA770080	HLA Class I motif
8	45	100.0	51	AA770080	Hepatitis C virus
9	45	100.0	53	AA770080	Anti-HCV antibody
10	45	100.0	62	AA770080	Anti-HCV antibody
11	45	100.0	62	AA770080	Anti-HCV antibody

12	45	100.0	65	AA770080	Hepatitis C virus
13	45	100.0	67	AA770080	Hepatitis C virus
14	45	100.0	71	AA770080	Anti-HCV antibody
15	45	100.0	71	AA770080	Anti-HCV antibody
16	45	100.0	72	AA770080	Anti-HCV antibody
17	45	100.0	75	AA770080	Anti-HCV antibody
18	45	100.0	75	AA770080	Anti-HCV antibody
19	45	100.0	77	AA770080	HCV NS-3 reactive
20	45	100.0	79	AA770080	Hepatitis C virus
21	45	100.0	80	AA770080	HCV antigenic NS3
22	45	100.0	81	AA770080	Anti-HCV antibody
23	45	100.0	81	AA770080	Anti-HCV antibody
24	45	100.0	81	AA770080	Anti-HCV antibody
25	45	100.0	81	AA770080	HCV NS-3 reactive
26	45	100.0	81	AA770080	HCV NS-3 reactive
27	45	100.0	81	AA770080	HCV NS-3 reactive
28	45	100.0	81	AA770080	HCV NS-3 reactive
29	45	100.0	81	AA770080	Prototype peptide
30	45	100.0	84	AA770080	Anti-HCV antibody
31	45	100.0	84	AA770080	Anti-HCV antibody
32	45	100.0	89	AA770080	Sequence encoded i
33	45	100.0	90	AA770080	Hepatitis C virus
34	45	100.0	93	AA770080	Anti-HCV antibody
35	45	100.0	102	AA770080	HCV c33c C-termina
36	45	100.0	159	AA770080	Polypeptide encode
37	45	100.0	159	AA770080	Hepatitis C virus
38	45	100.0	159	AA770080	Hepatitis C virus
39	45	100.0	159	AA770080	Hepatitis C virus
40	45	100.0	163	AA770080	Sequence of hepati
41	45	100.0	163	AA770080	Sequence encoded i
42	45	100.0	186	AA770080	Polypeptide encode
43	45	100.0	186	AA770080	Hepatitis C virus
44	45	100.0	186	AA770080	Hepatitis C virus
45	45	100.0	186	AA770080	Hepatitis C virus

ALIGNMENTS

RESULT 1

AA770080
ID AA770080 standard; peptide; 10 AA.

XX AA770080;

XX AC
XX 06-OCT-1995 (first entry)

XX JY-EBV cells pre-incubation peptide #1.

XX Hepatitis B virus polymerase; cytotoxic T cell response; prophylactic;
XX vaccine; chronic; acute HBV infection; carrier.

XX Hepatitis B virus.

XX WO9503777-A.

XX 09-FEB-1995.

XX 01-AUG-1994; 94WO-US08685.

XX 02-AUG-1993; 93US-0100870.

XX (SCRI) SCRIPPS RES INST.

XX Chisari FV;

XX WPI; 1995-082004/11.

XX New peptides inducing cytotoxic T lymphocytes to hepatitis B
XX virus - are regions of HB polymerase protein, for treating acute
XX and chronic infections

XX Example 2; Page 44; 85pp; English.

```

XX Peptides AAR70080-3 were pre-incubated with JY-EBV cells in an assay to
CC detect the cytotoxic activity of cytotoxic T cells (CTL), that were
CC challenged with hepatitis B virus (HBV) polymerase (HBpol) epitopes
CC e.g. the peptides GLSRVVARL or SGLSRVVARL, to the pre-incubated cells.
CC The peptide presented here is a hepatitis type C virus fragment.
CC The HBpol epitope peptides (see AAR70044-59) can be used,
CC prophylactically as vaccines, together with, or conjugated to, HBV helper
CC epitopes (AAR70060-4). The peptides can be used, particularly ex vivo,
CC to stimulate CTL cells, which cells can be reintroduced into patients who
CC have chronic or acute HBV infections or are carriers, especially in
CC treatments to prevent conversion from acute to chronic infections.
XX Sequence 10 AA;
SQ
Query Match 100.0%; Score 45; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVALGINAV 10
Db 1 klvalginav 10
IIIIIIIII
RESULT 2
AAR84574
ID AAR84574 standard; peptide; 10 AA.
XX
AC AAR84574;
XX
DT 25-APR-1996 (first entry)
XX
DE Cytotoxic T-cell epitope, aa 1406-1415 of HCV-1 NS3 region.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen.
XX
OS Hepatitis C virus.
XX
PN WO9525122-A1.
XX
PD 21-SEP-1995.
XX
PF 16-MAR-1995; 95WO-US03224.
XX
PR 17-MAR-1994; 94US-0214650.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Cerny A, Chisari FV;
XX
DR WPI; 1995-336941/43.
XX
PT Novel molecule comprising a cytotoxic T cell epitope - used to
PT vaccinate against hepatitis C viral infection
XX
PS Claim 1; Page 67; 85pp; English.
CC AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides
CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
CC were tested for peptide specific cytotoxic T-cell activity. The
CC peptides AAR84570-77 were found to have substantial homology with a T-
CC cell epitope and are useful in vaccines against HCV infection.
XX Sequence 10 AA;
SQ
Query Match 100.0%; Score 45; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVALGINAV 10
Db 1 klvalginav 10
IIIIIIIII

```

```

Db 1 klvalginav 10
RESULT 3
AAR84488
ID AAR84488 standard; peptide; 10 AA.
XX
AC AAR84488;
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus NS3 region epitopic peptide.
XX
KW Hepatitis C virus; HCV; immunogen; NS3 region; non-structural region;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.
XX
PS Claim 18; Page 68; 105pp; English.
XX
CC Peptides comprising at least 8 contiguous amino acids from the HCV
CC NS3 region between positions 1188 and 1463 and containing a
CC T-cell stimulating epitope are used in HCV immunogenic compositions.
CC The present sequence is a specific example of a T-cell epitope-
CC containing peptide which is preferred for use in the composition.
XX Sequence 10 AA;
SQ
Query Match 100.0%; Score 45; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVALGINAV 10
Db 1 klvalginav 10
IIIIIIIII
RESULT 4
AAW39560
ID AAW39560 standard; peptide; 10 AA.
XX
AC AAW39560;
XX
DT 11-JUN-1998 (first entry)
XX
DE HCV1 NS3 peptide (pos.1406-1415) capable of binding to HLA-A*0201.
XX
KW T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW vaccine; immunogenic; major histocompatibility complex; MHC; B cell;
KW disease; anti-tumour; anti-viral.
XX
OS Hepatitis C virus.
XX
PN WO9741440-A1.

```


XX PD 06-NOV-1997.
 XX XX
 XX PF 28-APR-1997; 97WO-NL00229.
 XX PR 23-DEC-1996; 96EP-0203670.
 XX PR 26-APR-1996; 96EP-0201145.
 XX XX
 XX (UYLE-) RIJKSUNIV LEIDEN.
 XX PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
 XX PI Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;
 XX XX WPI; 1997-549891/50.
 XX DR
 XX XX Method of selecting T cell peptide epitope(s) - by measuring the
 XX PT stability of HLA class I-peptide complexes on intact B cells
 XX PT
 XX PS Example 2; Page 66; 109pp; English.
 XX XX
 XX CC Peptides AAW39430-W39734 are used in a novel method for the selection of
 XX CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
 XX CC method involves the identification of peptide sequences capable of
 XX CC binding to an HLA (human leukocyte antigen) class I molecule and
 XX CC measuring the binding of this epitope peptide to the HLA class I
 XX CC peptide. The stability of binding of the peptide and MHC (major
 XX CC histocompatibility complex) class I molecule is measured on intact human
 XX CC B cells carrying the MHC molecule at their cell surfaces. The method can
 XX CC be used to select peptide epitopes for generating vaccines against a
 XX CC disease associated with the polypeptide, e.g. cancers or AIDS. The
 XX CC peptide epitopes are especially T-cell peptide epitopes with strong
 XX CC anti-tumour and anti-viral immune responses. Peptide AAW39560 is a
 XX CC conserved hepatitis C virus type 1 NS3 derived peptide used to determine
 XX CC the stability of HLA-A*0201 complexed with known CTL lymphocytes.
 XX XX
 XX SQ Sequence 10 AA;
 XX
 Query Match 100.0%; Score 45; DB 18; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLVALGINAV 10
 Db | | | | | | | | | |
 1 klvalginav 10
 RESULT 5
 AAW54634
 ID AAW54634 standard; peptide; 10 AA.
 XX AC
 XX AC AAW54634;
 XX DT 25-SEP-1998 (first entry)
 XX DE Peptide from HCV-1 1406-1415.
 XX XX
 XX KW Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell;
 XX KW vaccine; treatment.
 XX OS Synthetic.
 XX XX
 XX PN WO9813378-A1.
 XX XX
 XX PD 02-APR-1998.
 XX XX
 XX PF 25-SEP-1997; 97WO-NL00536.
 XX XX
 XX PR 26-SEP-1996; 96EP-0202701.
 XX XX
 XX (UYLE-) RIJKSUNIV LEIDEN.
 XX PA
 XX PI Drijfhout JW, Koning F;

XX WPI; 1998-230631/20.
 XX DR
 XX XX Increasing uptake and presentation of antigen(s) - by adding mannose
 XX PT residue(s) to antigen for increasing T cell response, useful in,
 XX PT e.g. vaccines against viral infection(s)
 XX PS
 XX PS Disclosure; Page 28; 47pp; English.
 XX XX
 XX CC The peptides AAW54559-W54809 are examples of peptides to which at least
 XX CC 1 (preferably 2) mannose can be attached to increase their uptake as
 XX CC antigens by antigen-presenting cells. Uptake of agonist mannosylated
 XX CC peptides will increase the T cell response, whereas uptake of antagonist
 XX CC peptides blocks the T cell response. Blocking binding of immunogenic
 XX CC autoantigens can be used in treatment of type I diabetes, rheumatoid
 XX CC arthritis, graft rejection etc., also to induce T-cell non-
 XX CC responsiveness. Vaccines containing mannosylated antigen are used to
 XX CC prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths
 XX CC and parasites.
 XX XX
 XX SQ Sequence 10 AA;
 XX
 Query Match 100.0%; Score 45; DB 19; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLVALGINAV 10
 Db | | | | | | | | | |
 1 klvalginav 10
 RESULT 6
 AAY10138
 ID AAY10138 standard; Peptide; 10 AA.
 XX AC
 XX AC AAY10138;
 XX DT 12-MAY-1999 (first entry)
 XX XX
 XX DE T cell epitope/MHC ligand SEQ ID NO:68.
 XX XX
 XX KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 XX KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 XX KW malignant melanoma; viral disease; hepatitis; AIDS.
 XX OS Synthetic.
 XX OS Hepatitis C virus.
 XX PN WO9902183-A2.
 XX XX
 XX PD 21-JAN-1999.
 XX XX
 XX PF 10-JUL-1998; 98WO-US14289.
 XX XX
 XX PR 10-DEC-1997; 97US-0988320.
 XX PR 10-JUL-1997; 97CA-2209815.
 XX XX
 XX PA (CTLI-) CTL IMMUNOTHERAPIES CORP.
 XX XX
 XX PI Kuendig TM, Simard JUL;
 XX XX
 XX DR WPI; 1999-120514/10.
 XX XX
 XX XX Inducing a cytotoxic T lymphocyte response - by maintaining a level
 XX PT of antigen in the lymphatic system of a mammal so as to provide a
 XX PT sustained CTL response, used to treat, e.g. AIDS
 XX XX
 XX PS Disclosure; Page 25; 199pp; English.
 XX XX
 XX CC The present invention describes a method of inducing and/or sustaining
 XX CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 XX CC method comprises: (a) delivering an antigen to the mammal at a level to

CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX
 SQ Sequence 10 AA;

Query Match 100.0%; Score 45; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
 Db 1 klvalginav 10
 |||||

RESULT 7
 AAY10440
 ID AAY10440 standard; Peptide; 10 AA.
 XX
 AC AAY10440;
 XX
 DT 12-MAY-1999 (first entry)
 XX
 DE HLA Class I motif peptide SEQ ID NO:370.
 XX
 KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

OS Synthetic.
 OS Hepatitis C virus.
 XX
 PN WO9902183-A2.
 XX
 PD 21-JAN-1999.
 XX
 PF 10-JUL-1998; 98WO-US14289.
 XX
 PR 10-DEC-1997; 97US-0988320.
 PR 10-JUL-1997; 97CA-2209815.

XX (CTLI-) CTL IMMUNOTHERAPIES CORP.
 XX
 PI Kuendig TM, Simard JLL;
 XX
 DR WPI; 1999-120514/10.
 XX
 PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX
 PS Disclosure; Page 40; 199pp; English.

CC The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of

CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

SQ Sequence 10 AA;

Query Match 100.0%; Score 45; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
 Db 1 klvalginav 10
 |||||

RESULT 8
 AAR59858
 ID AAR59858 standard; Protein; 51 AA.
 XX
 AC AAR59858;
 XX
 DT 23-JAN-1995 (first entry)
 XX
 DE Hepatitis C virus NS-3 antigen fragment.
 XX
 KW Hepatitis C virus; NS-3 antigen; diagnostic.

OS Escherichia coli JM101.
 XX
 PN WO9413700-A.
 XX
 PD 23-JUN-1994.
 XX
 PF 07-DEC-1993; 93WO-EP03478.

XX
 PR 07-DEC-1992; 92EP-0203802.
 PR 25-JUN-1993; 93EP-0201854.
 XX
 PA (ALKU) AKZO NOBEL NV.
 XX
 PI Boender PJ, Habets WJA;
 XX
 DR WPI; 1994-234210/28.
 DR N-PSDB; AAQ69124.

XX New peptide(s) corresponding to hepatitis C virus - used for
 PT detection of hepatitis C virus antibodies and treatment or
 PT prevention of infection or for prodn. of antibodies
 XX
 PS Claim 2; Page 46; 52pp; English.
 XX
 CC The HCV NS-3 region peptide is highly immunoreactive with HCV
 CC antibodies and can provide highly sensitive and specific assays
 CC for diagnosis and monitoring. It can also be used for
 CC prevention/treatment of non-A non-B hepatitis.

XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 45; DB 15; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.056;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
 Db 3 klvalginav 12
 |||||

RESULT 9

AAR69538
ID AAR69538 standard; peptide; 53 AA.
XX AC AAR69538;
XX DT 05-SEP-1995 (first entry)
XX DE Anti-HCV antibody immunoreactive PepB peptide L4A.
XX DE Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;
KW immunoreactive peptide; PepB; L4A; infection detection assay.
XX OS Synthetic.
XX PN WO9500670-A.
XX PD 05-JAN-1995.
XX PF 22-JUN-1994; 94WO-US07088.
XX PR 28-JUN-1993; 93US-0083947.
XX PA (UNBI-) UNITED BIOMEDICAL INC.
XX PI Hosein B, Wang CY;
XX DR WPI; 1995-052105/07.
XX PT Linear and branched peptide(s) comprising, e.g. PepB and PepC -
PT useful in vaccines and assays for non-A, non-B hepatitis
XX Claim 8; Page 45; 58pp; English.

AAR69523-R69545 are anti-hepatitis C virus (HCV) antibody
immunoreactive peptides, based on the PepB (AAR69529) or the PepC
(AAR69534) sequences. These peptides, in linear or
branched dimer compsns. are used in immunoassays for the
detection of HCV infections, the peptide compsns. may also be
useful in vaccines against these infections.
XX Sequence 53 AA;

Query Match 100.0%; Score 45; DB 16; Length 53;
Best Local Similarity 100.0%; Pred. NO. 0.058;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
| | | | |
Db 10 klvalginav 19

RESULT 10

AAR69526
ID AAR69526 standard; peptide; 62 AA.
XX AC AAR69526;
XX DT 05-SEP-1995 (first entry)
XX DE Anti-HCV antibody immunoreactive PepB peptide L1A.
XX DE Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;
KW immunoreactive peptide; PepB; L1A; infection detection assay.
XX OS Synthetic.
XX PN WO9500670-A.
XX PD 05-JAN-1995.
XX PF 22-JUN-1994; 94WO-US07088.

XX 28-JUN-1993; 93US-0083947.
XX (UNBI-) UNITED BIOMEDICAL INC.
XX PI Hosein B, Wang CY;
XX DR WPI; 1995-052105/07.
XX PT Linear and branched peptide(s) comprising, e.g. PepB and PepC -
PT useful in vaccines and assays for non-A, non-B hepatitis
XX Claim 8; Pages 38-39; 58pp; English.
XX AAR69523-R69545 are anti-hepatitis C virus (HCV) antibody
immunoreactive peptides, based on the PepB (AAR69529) or the PepC
(AAR69534) sequences. These peptides, in linear or
branched dimer compsns. are used in immunoassays for the
detection of HCV infections, the peptide compsns. may also be
useful in vaccines against these infections.
XX Sequence 62 AA;

Query Match 100.0%; Score 45; DB 16; Length 62;
Best Local Similarity 100.0%; Pred. NO. 0.069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
| | | | |
Db 10 klvalginav 19

RESULT 11

AAR69539
ID AAR69539 standard; peptide; 62 AA.
XX AC AAR69539;
XX DT 05-SEP-1995 (first entry)
XX DE Anti-HCV antibody immunoreactive PepB peptide L4B.
XX DE Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;
KW immunoreactive peptide; PepB; L4B; infection detection assay.
XX OS Synthetic.
XX PN WO9500670-A.
XX PD 05-JAN-1995.
XX PF 22-JUN-1994; 94WO-US07088.
XX PR 28-JUN-1993; 93US-0083947.
XX PA (UNBI-) UNITED BIOMEDICAL INC.
XX PI Hosein B, Wang CY;
XX DR WPI; 1995-052105/07.
XX PT Linear and branched peptide(s) comprising, e.g. PepB and PepC -
XX useful in vaccines and assays for non-A, non-B hepatitis
XX Claim 8; Page 46; 58pp; English.
XX AAR69523-R69545 are anti-hepatitis C virus (HCV) antibody
immunoreactive peptides, based on the PepB (AAR69529) or the PepC
(AAR69534) sequences. These peptides, in linear or
branched dimer compsns. are used in immunoassays for the
detection of HCV infections, the peptide compsns. may also be
useful in vaccines against these infections.

```

XX SQ Sequence 62 AA;
Query Match 100.0%; Score 45; DB 16; Length 62;
Best Local Similarity 100.0%; Pred. No. 0.069; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
DB 19 klvalginav 28

RESULT 12
AAR59857
ID AAR59857 standard; Protein; 65 AA.
AC AAR59857;
XX
DT 23-JAN-1995 (first entry)
XX
DE Hepatitis C virus NS-3 antigen fragment.
XX
KW Hepatitis C virus; NS-3 antigen; diagnostic.
XX
OS Escherichia coli JM101.
XX
PN WO9413700-A.
XX
PD 23-JUN-1994.
XX
PF 07-DEC-1993; 93WO-EP03478.
XX
PR 07-DEC-1992; 92EP-0203802.
XX
PR 25-JUN-1993; 93EP-0201854.
XX
PA (ALKU ) AKZO NOBEL NV.
XX
PI Boender PJ, Habets WJA;
XX
DR WPI; 1994-234210/28.
XX
DR N-PSDB; AAQ69123.
XX
PT New peptide(s) corresponding to hepatitis C virus - used for
PT detection of hepatitis C virus antibodies and treatment or
PT prevention of infection or for prodn. of antibodies
XX
PS Claim 2; Page 44-45; 52pp; English.
XX
CC The HCV NS-3 region peptide is highly immunoreactive with HCV
CC antibodies and can provide highly sensitive and specific assays
CC for diagnosis and monitoring. It can also be used for
CC prevention/treatment of non-A non-B hepatitis.
XX
SQ Sequence 67 AA;

Query Match 100.0%; Score 45; DB 15; Length 67;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
DB 19 klvalginav 28

RESULT 14
AAR69524
ID AAR69524 standard; peptide; 71 AA.
XX
AC AAR69524;
XX
DT 05-SEP-1995 (first entry)
XX
DE Anti-HCV antibody immunoreactive PepB peptide C13.
XX
KW Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;
KW immunoreactive peptide; PepB; C13; infection detection assay.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 6
FT /label= Nva
XX
PN WO9500670-A.
XX
PD 05-JAN-1995.
XX
PF 22-JUN-1994; 94WO-US07088.
XX
PR 28-JUN-1993; 93US-0083947.
XX

```

```

XX DE Hepatitis C virus NS-3 antigen fragment.
XX
KW Hepatitis C virus; NS-3 antigen; diagnostic.
XX
OS Escherichia coli JM101.
XX
PN WO9413700-A.
XX
PD 23-JUN-1994.
XX
PF 07-DEC-1993; 93WO-EP03478.
XX
PR 07-DEC-1992; 92EP-0203802.
XX
PR 25-JUN-1993; 93EP-0201854.
XX
PA (ALKU ) AKZO NOBEL NV.
XX
PI Boender PJ, Habets WJA;
XX
DR WPI; 1994-234210/28.
XX
DR N-PSDB; AAQ69122.
XX
PT New peptide(s) corresponding to hepatitis C virus - used for
PT detection of hepatitis C virus antibodies and treatment or
PT prevention of infection or for prodn. of antibodies
XX
PS Claim 2; Page 44-45; 52pp; English.
XX
CC The HCV NS-3 region peptide is highly immunoreactive with HCV
CC antibodies and can provide highly sensitive and specific assays
CC for diagnosis and monitoring. It can also be used for
CC prevention/treatment of non-A non-B hepatitis.
XX
SQ Sequence 67 AA;

Query Match 100.0%; Score 45; DB 15; Length 67;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
DB 19 klvalginav 28

RESULT 14
AAR69524
ID AAR69524 standard; peptide; 71 AA.
XX
AC AAR69524;
XX
DT 05-SEP-1995 (first entry)
XX
DE Anti-HCV antibody immunoreactive PepB peptide C13.
XX
KW Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;
KW immunoreactive peptide; PepB; C13; infection detection assay.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 6
FT /label= Nva
XX
PN WO9500670-A.
XX
PD 05-JAN-1995.
XX
PF 22-JUN-1994; 94WO-US07088.
XX
PR 28-JUN-1993; 93US-0083947.
XX

```

PA (UNBI-) UNITED BIOMEDICAL INC.

XX Hosein B, Wang CY;

XX WPI; 1995-052105/07.

XX Linear and branched peptide(s) comprising, e.g. PepB and PepC -
PT useful in vaccines and assays for non-A, non-B hepatitis

XX Claim 8; Pages 36-37; 58pp; English.

XX AAR69523-R69545 are anti-hepatitis C virus (HCV) antibody
CC immunoreactive peptides, based on the PepB (AAR69529) or the PepC
CC (AAR69534) sequences. These peptides, in linear or
CC branched dimer compsns. are used in immunoassays for the
CC detection of HCV infections, the peptide compsns. may also be
CC useful in vaccines against these infections.

XX Sequence 71 AA;

Query Match 100.0%; Score 45; DB 16; Length 71;

Best Local Similarity 100.0%; Pred. No. 0.08; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10

Db 19 klvalginav 28

RESULT 15

AAR69527

ID AAR69527 standard; peptide; 71 AA.

XX AC AAR69527;

XX DT 05-SEP-1995 (first entry)

XX Anti-HCV antibody immunoreactive PepB peptide LIB.

XX Hepatitis C virus; HCV; non-A non-B; antibodies; vaccines;

XX immunoreactive peptide; PepB; LIB; infection detection assay.

XX Synthetic.

XX WO9500670-A.

XX PD 05-JAN-1995.

XX PF 22-JUN-1994; 94WO-US07088.

XX PR 28-JUN-1993; 93US-0083947.

XX PA (UNBI-) UNITED BIOMEDICAL INC.

XX Hosein B, Wang CY;

XX WPI; 1995-052105/07.

XX Linear and branched peptide(s) comprising, e.g. PepB and PepC -
PT useful in vaccines and assays for non-A, non-B hepatitis

XX Claim 8; Page 39; 58pp; English.

XX AAR69523-R69545 are anti-hepatitis C virus (HCV) antibody
CC immunoreactive peptides, based on the PepB (AAR69529) or the PepC
CC (AAR69534) sequences. These peptides, in linear or
CC branched dimer compsns. are used in immunoassays for the
CC detection of HCV infections, the peptide compsns. may also be
CC useful in vaccines against these infections.

XX Sequence 71 AA;

Query Match 100.0%; Score 45; DB 16; Length 71;

Best Local Similarity 100.0%; Pred. No. 0.08; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10

Db 19 klvalginav 28

Search completed: August 23, 2002, 10:01:00

Job time: 390 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:52 ; Search time 61.77 Seconds
(without alignments)
3.954 Million cell updates/sec

Title: US-08-854-825-28

Perfect score: 45

Sequence: 1 KLVALGINAV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	10	1	US-08-214-650-28
2	45	100.0	53	1	US-08-083-947-16
3	45	100.0	53	5	PCT-US94-07088-16
4	45	100.0	62	1	US-08-083-947-4
5	45	100.0	62	1	US-08-083-947-17
6	45	100.0	62	5	PCT-US94-07088-4
7	45	100.0	62	5	PCT-US94-07088-17
8	45	100.0	71	1	US-08-083-947-2
9	45	100.0	71	1	US-08-083-947-5
10	45	100.0	71	5	PCT-US94-07088-2
11	45	100.0	71	5	PCT-US94-07088-5
12	45	100.0	72	1	US-08-083-947-18
13	45	100.0	72	5	PCT-US94-07088-18
14	45	100.0	75	1	US-08-083-947-19
15	45	100.0	75	1	US-08-530-550-5
16	45	100.0	75	5	PCT-US94-07088-19
17	45	100.0	75	5	PCT-US95-13660-5
18	45	100.0	77	1	US-08-530-550-6
19	45	100.0	77	5	PCT-US95-13660-6
20	45	100.0	81	1	US-08-083-947-1
21	45	100.0	81	1	US-08-083-947-6
22	45	100.0	81	1	US-08-530-550-4
23	45	100.0	81	1	US-08-530-550-10
24	45	100.0	81	1	US-08-530-550-12
25	45	100.0	81	1	US-08-530-550-13
26	45	100.0	81	1	US-08-530-550-15
27	45	100.0	81	5	PCT-US94-07088-1

28	45	100.0	81	5	PCT-US94-07088-6	Sequence 6, Appl
29	45	100.0	81	5	PCT-US95-13660-4	Sequence 4, Appl
30	45	100.0	81	5	PCT-US95-13660-10	Sequence 10, Appl
31	45	100.0	81	5	PCT-US95-13660-12	Sequence 12, Appl
32	45	100.0	81	5	PCT-US95-13660-13	Sequence 13, Appl
33	45	100.0	81	5	PCT-US95-13660-15	Sequence 15, Appl
34	45	100.0	84	1	US-08-083-947-8	Sequence 8, Appl
35	45	100.0	84	1	US-08-530-550-21	Sequence 21, Appl
36	45	100.0	84	5	PCT-US94-07088-8	Sequence 8, Appl
37	45	100.0	84	5	PCT-US95-13660-21	Sequence 21, Appl
38	45	100.0	89	4	US-08-444-818-22	Sequence 22, Appl
39	45	100.0	93	1	US-08-083-947-7	Sequence 7, Appl
40	45	100.0	93	5	PCT-US94-07088-7	Sequence 7, Appl
41	45	100.0	102	1	US-08-411-913-9	Sequence 9, Appl
42	45	100.0	159	1	US-07-853-985A-8	Sequence 8, Appl
43	45	100.0	159	1	US-07-681-703B-8	Sequence 8, Appl
44	45	100.0	159	1	US-08-184-236-8	Sequence 8, Appl
45	45	100.0	159	2	US-08-407-410B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-28
; Sequence 28, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-214-650-28

Query Match 100.0%; Score 45; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
|||||

; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-083-947-4

Query Match 100.0%; Score 45; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 10 KLVALGINAV 19

RESULT 5
US-083-947-17
; Sequence 17, Application US/08083947
; Patent No. 5639594
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
; TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-B Hepat
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,947
; FILING DATE: 19930628
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 946,054
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Willson, M. Lisa
; REGISTRATION NUMBER: 34045
; REFERENCE/DOCKET NUMBER: 20002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-083-947-17

Query Match 100.0%; Score 45; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 19 KLVALGINAV 28

RESULT 6
PCT-US94-07088-4
; Sequence 4, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PCI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US94-07088-4

Query Match 100.0%; Score 45; DB 5; Length 62;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 10 KLVALGINAV 19

RESULT 7
PCT-US94-07088-17
; Sequence 17, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA

ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07088
FILING DATE: 22-JUNE-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,947
FILING DATE: 28-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: LIN, MARIA C.H.
REGISTRATION NUMBER: 29323
REFERENCE/DOCKET NUMBER: 1151-4101PC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 62 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US94-07088-17

Query Match 100.0%; Score 45; DB 5; Length 62;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
DB 19 KLVALGINAV 28

RESULT 8
US-08-083-947-2
Sequence 2, Application US/08083947
Patent No. 5639594
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
APPLICANT: Hosein, Barbara
TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-B Hepat
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson
STREET: 25 Davids Drive
CITY: Hauppauge
STATE: NY
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,947
FILING DATE: 19930628
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 946,054
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34045
REFERENCE/DOCKET NUMBER: 20002
TELEPHONE: (516)273-2828

TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 71 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note= "NORVALINE"
US-08-083-947-2

Query Match 100.0%; Score 45; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
DB 19 KLVALGINAV 28

RESULT 9
US-08-083-947-5
Sequence 5, Application US/08083947
Patent No. 5639594
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
APPLICANT: Hosein, Barbara
TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-B He
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson
STREET: 25 Davids Drive
CITY: Hauppauge
STATE: NY
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,947
FILING DATE: 19930628
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 946,054
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34045
REFERENCE/DOCKET NUMBER: 20002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 71 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-083-947-5

Query Match 100.0%; Score 45; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10

Db 19 KLVALGINAV 28
|||||

RESULT 10

PCT-US94-07088-2

; Sequence 2, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 6
; OTHER INFORMATION: /note= "NORVALINE"
PCT-US94-07088-2

Query Match 100.0%; Score 45; DB 5; Length 71;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 19 KLVALGINAV 28

RESULT 11

PCT-US94-07088-5

; Sequence 5, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis

; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US94-07088-5

Query Match 100.0%; Score 45; DB 5; Length 71;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 19 KLVALGINAV 28

RESULT 12

US-08-083-947-18

; Sequence 18, Application US/08083947
; Patent No. 5639594
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
; TITLE OF INVENTION: In Diagnosing and Detecting No. 5639594-A, No. 5639594-B He
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,947
; FILING DATE: 19930628
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 946,054

; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34045
; REFERENCE/DOCKET NUMBER: 2000Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-083-947-18

Query Match 100.0%; Score 45; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 29 KLVALGINAV 38

RESULT 13
PCT-US94-07088-18
; Sequence 18, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PCI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US94-07088-18

Query Match 100.0%; Score 45; DB 5; Length 72;

Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 29 KLVALGINAV 38

RESULT 14
US-08-083-947-19
; Sequence 19, Application US/08083947
; Patent No. 5639594
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
; TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-B He
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,947
; FILING DATE: 19930628
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 946,054
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34045
; REFERENCE/DOCKET NUMBER: 2000Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-083-947-19

Query Match 100.0%; Score 45; DB 1; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 32 KLVALGINAV 41

RESULT 15
US-08-530-550-5
; Sequence 5, Application US/08530550
; Patent No. 5736321
; GENERAL INFORMATION:
; APPLICANT: Hosein, Barbara
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Peptides Effective for Diagnosis and
; TITLE OF INVENTION: Detection of Hepatitis c Infection
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:

Search completed: August 23, 2002, 09:57:52
Job time: 202 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:47 ; Search time 78.82 Seconds
(without alignments)
12.191 Million cell updates/sec

Title: US-08-854-825-28

Perfect score: 45

Sequence: 1 KLVALGINAV 10

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	386	2 S68016	ATPase/RNA helicase
2	45	100.0	3011	1 GNVWC3	genome polyprotein
3	45	100.0	3011	1 GNVWCH	genome polyprotein
4	44	97.8	3011	1 S40770	genome polyprotein
5	38	84.4	3010	1 S18030	genome polyprotein
6	37	82.2	216	2 S21337	genome polyprotein
7	36	80.0	595	2 T29434	beta-galactosidase
8	35	77.8	229	2 A69781	conserved hypotet
9	35	77.8	282	2 AB0828	probable transcrip
10	35	77.8	3010	1 GNVWTC	genome polyprotein
11	35	77.8	3010	1 GNVWCJ	genome polyprotein
12	35	77.8	3010	1 A45573	genome polyprotein
13	34	75.6	124	2 B69192	conserved hypotet
14	34	75.6	342	2 JC4092	DMC1/LIM15 homolog
15	34	75.6	345	2 T08838	RecA/Rad51/DMC1-11
16	34	75.6	432	2 AI2097	phosphate permease
17	34	75.6	3010	1 GNVWTW	genome polyprotein
18	33	73.3	310	2 E64751	probable membrane
19	33	73.3	402	2 D72400	hypothetical prote
20	33	73.3	876	2 PC2219	polypeptide - hepa
21	32	71.1	62	2 F69015	phosphoenolpyruvat
22	32	71.1	120	2 B89999	conserved hypotet
23	32	71.1	194	2 JS0864	interferon precurs
24	32	71.1	248	2 AI0702	probable ABC trans
25	32	71.1	280	2 T03559	2-hydroxyhepta-2,4
26	32	71.1	336	2 T30504	hypothetical prote
27	32	71.1	345	2 C95873	probable malate de
28	32	71.1	353	2 H75446	(S)-2-hydroxy-acid
29	32	71.1	440	2 H97339	probable permease

30	32	71.1	445	2 D83819	hypothetical prote
31	32	71.1	468	2 C82722	UDP-N-acetylmutamo
32	32	71.1	499	2 S07640	glutelin precursor
33	32	71.1	499	2 S17762	glutelin glub-1 pr
34	32	71.1	500	2 A87302	amino acid permeas
35	32	71.1	509	2 G82809	t-complex protein
36	32	71.1	568	2 S56779	heat shock protein
37	32	71.1	679	2 S21764	protein kinase STE
38	32	71.1	738	1 S51380	hypothetical prote
39	32	71.1	2163	2 T15276	hypothetical prote
40	31	68.9	74	2 H87490	conserved hypotet
41	31	68.9	106	2 G82395	probable RNasep-as
42	31	68.9	167	2 A85023	hypothetical prote
43	31	68.9	193	2 A11904	hypothetical prote
44	31	68.9	208	2 G81018	hypothetical prote
45	31	68.9	214	2 S39831	hypothetical prote

ALIGNMENTS

RESULT 1

S68016

ATPase/RNA helicase - hepatitis C virus (fragment)

C:Species: hepatitis C virus

C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 02-Feb-2001

C:Accession: S68016

R:Jin, L.; Peterson, D.L.

Arch. Biochem. Biophys. 323, 47-53, 1995

A:Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/

A:Reference number: S68016; MUID:96019946

A:Accession: S68016

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-386 <JIN>

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polyprotein

F:24-31/Region: nucleotide-binding motif A (P-loop)

F:86-91/Region: nucleotide-binding motif B

F:90-93/Region: DEXH motif

Query Match 100.0%; Score 45; DB 2; Length 386;

Best Local Similarity 100.0%; Pred. No. 0.3;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10

Db 180 KLVALGINAV 189

RESULT 2

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>
 A:Cross-references: DDBJ:D10128
 A:Experimental source: isolates E-b16
 A:Accession: PQ0404
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein M #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1318-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
 |||||
 DB 1406 KLVALGINAV 1415

RESULT 3
 genome polyprotein - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA5534.1; PID:g329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar
 A:Reference number: A41546; MUID:92052256
 A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein M #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1318-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
 |||||
 DB 1406 KLVALGINAV 1415

RESULT 4
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PC1285
 R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <ORA>
 A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BA01582.1; PID:g221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PC1284; MUID:91013116
 A:Accession: PC1285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein M #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 97.8%; Score 44; DB 1; Length 3011;
 Best Local Similarity 90.0%; Pred. No. 3.3;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
 |||||
 DB 1406 KLVALGINAV 1415

RESULT 5
 genome polyprotein - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
 A:Reference number: S18028
 A:Accession: S18030

Query Match 97.8%; Score 44; DB 1; Length 3011;
 Best Local Similarity 90.0%; Pred. No. 3.3;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
 |||||
 DB 1406 KLVALGINAV 1415

RESULT 5
 genome polyprotein - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
 A:Reference number: S18028
 A:Accession: S18030

AB0620
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence revision 09-Nov-2001 #text change 09-Nov-2001
probable transcription regulator yfH [imported] - Salmonella enterica subsp. enteric

AB0620
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: This species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence revision 09-Nov-2001 #text change 09-Nov-2001
probable transcription regulator yfH [imported] - Salmonella enterica subsp. enteric

C:Accession: AB0828
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AB0828
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-282 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD02774.1; PID:gl6503784; GSPDB:GN00176
 C:Genetics:
 A:Gene: yfhH

Query Match 77.8%; Score 35; DB 2; Length 282;
 Best Local Similarity 70.0%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KLVALGINAV 10
 DB 153 KLLKIGINAV 162
 ||:|||||
 RESULT 10
 GNMVTC
 genome polypeptide - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C:Species: hepatitis C virus
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 C:Accession: A38465
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
 J. Virol. 65, 1105-1113, 1991
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human
 A:Reference number: A38465; MUID:91140698
 A:Accession: A38465
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <TAK>
 A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus (strain J)
 F:1230-1237/Product: hepatitis C virus (strain J)
 F:1312-1317/Product: hepatitis C virus (strain J)
 F:1316-1319/Product: hepatitis C virus (strain J)
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 77.8%; Score 35; DB 1; Length 3010;
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KLVALGINAV 10
 DB 1406 KLSGLGINAV 1415
 |||||||
 RESULT 11
 GNMVCT
 genome polypeptide - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Y.; Shimotohno, K.
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier
 A:Reference number: A45573; MUID:92295714
 A:Accession: A45573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <TAN>
 A:Cross-references: GB:D11168; GB:D011171; NID:g221612; PIDN:BAA01943.1; PID:g221613
 A:Experimental source: HCV-JT
 A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus (strain J)
 F:1230-1237/Product: hepatitis C virus (strain J)
 F:1312-1317/Product: hepatitis C virus (strain J)
 F:1316-1319/Product: hepatitis C virus (strain J)
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
 C:Accession: A39253; PS0086
 R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi
 Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient
 A:Reference number: A39253; MUID:91088550
 A:Accession: A39253
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <KAT>
 A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
 R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
 Proc. Jpn. Acad. 65B, 219-223, 1989
 A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v
 A:Reference number: PS0085
 A:Accession: PS0086
 A:Molecule type: genomic RNA
 A:Residues: 2650-2707 <KA2>
 A:Experimental source: Japanese isolate
 C:Comment: The cleavage sites of this polypeptide have not been determined.
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus (strain J)
 F:1230-1237/Product: hepatitis C virus (strain J)
 F:1312-1317/Product: hepatitis C virus (strain J)
 F:1316-1319/Product: hepatitis C virus (strain J)
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224

Query Match 77.8%; Score 35; DB 1; Length 3010;
 Best Local Similarity 70.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KLVALGINAV 10
 DB 1406 KLTGLGINAV 1415
 |||||
 RESULT 12
 A45573
 genome polypeptide - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J)
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Y.; Shimotohno, K.
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier
 A:Reference number: A45573; MUID:92295714
 A:Accession: A45573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <TAN>
 A:Cross-references: GB:D11168; GB:D011171; NID:g221612; PIDN:BAA01943.1; PID:g221613
 A:Experimental source: HCV-JT
 A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus (strain J)
 F:1230-1237/Product: hepatitis C virus (strain J)
 F:1312-1317/Product: hepatitis C virus (strain J)
 F:1316-1319/Product: hepatitis C virus (strain J)
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224

F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 77.8%; Score 35; DB 1; Length 3010;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
II:IIIIII
DB 1406 KLSGLGINAV 1415

RESULT 13

B69192 conserved hypothetical protein MTH691 - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: B69192
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Kt, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional genome
A:Reference number: A69000; MUID:98037514
A:Accession: B69192

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-124 <MTH>
A:Cross-references: GB:AE000848; GB:AE000666; NID:g2621761; PIDN:AAB85196.1; PID:g262177
A:Experimental source: Strain Delta H
C:Genetics:
A:Gene: MTH691
C:Superfamily: conserved hypothetical protein MJ0742

Query Match 75.6%; Score 34; DB 2; Length 124;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLVALGINA 9
II:IIIIII
DB 61 KLIALGITA 69

RESULT 14

JC4092 DMCL/LIM15 homolog - Arabidopsis thaliana
N:Alternate names: recA-like protein
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 02-Feb-2001
C:Accession: JC4092
R:Sato, S.; Hotta, Y.; Tabata, S.
DNA Res. 2, 89-93, 1995
A:Title: Structural analysis of a recA-like gene in the genome of Arabidopsis thaliana.
A:Reference number: JC4092; MUID:96093929
A:Accession: JC4092
A:Molecule type: mRNA
A:Residues: 1-342 <SAT>
A:Cross-references: DBJ:D45415; NID:g871831; PIDN:BAA08255.1; PID:g871832
C:Comment: This protein plays key roles in genetic recombination by finding homologous sequences
C:Genetics:
A:Gene: ArLIM15
A:Introns: 6/1; 34/2; 64/1; 88/1; 101/1; 121/2; 136/1; 151/3; 172/1; 194/3; 225/3; 248/3
C:Superfamily: yeast DNA repair protein RAD51
C:Keywords: nucleotide binding; P-loop
F:133-140/Region: nucleotide-binding motif A (P-loop)
F:224-229/Region: nucleotide-binding motif B

Query Match 75.6%; Score 34; DB 2; Length 342;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLVALGINA 9
II:IIIIII
DB 33 KLIAOGINA 41

RESULT 15

T08838 RecA/Rad51/DMC1-like protein - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08838
R:Hadi, M.Z.; Finer, J.J.
submitted to the EMBL Data Library, August 1996
A:Description: RecA like gene from soybean.
A:Reference number: Z16485
A:Accession: T08838
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA
A:Residues: 1-345 <HAD>
A:Cross-references: EMBL:U66836; NID:g1518156; PID:g1518157
C:Function:
A:Description: recombinase
C:Superfamily: yeast DNA repair protein RAD51

Query Match 75.6%; Score 34; DB 2; Length 345;
Best Local Similarity 77.8%; Pred. No. 44;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLVALGINA 9
II:IIIIII
DB 34 KLIAOGINA 42

Search completed: August 23, 2002, 09:56:49
Job time: 139 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:22 ; Search time 37 Seconds
(without alignments)
10.465 Million cell updates/sec

Title: US-08-854-825-28
Perfect score: 45
Sequence: 1 KLVALGINAV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	45	100.0	3011	1 POLG_HCVH	P27958 h genome po
3	35	77.8	701	1 HS90_PODAN	O43109 podospora a
4	35	77.8	3010	1 POLG_HCVBK	P26663 h genome po
5	35	77.8	3010	1 POLG_HCVJA	P26662 h genome po
6	35	77.8	3010	1 POLG_HCVJT	Q00269 h genome po
7	34	75.6	342	1 DMCL_ARATH	Q39009 arabidopsis
8	34	75.6	345	1 DMCL_SOYBN	Q96449 glycine max
9	34	75.6	3010	1 POLG_HCVTW	P29846 h genome po
10	33	73.3	310	1 MMUM_ECOLI	O47690 escherichia
11	32	71.1	194	1 INA_FELCA	P35849 felis silve
12	32	71.1	441	1 HS82_ASPPU	P40292 aspergillus
13	32	71.1	499	1 GLUA_ORVSA	P14323 oryza sativ
14	32	71.1	499	1 GLUB_ORVSA	Q02898 oryza sativ
15	32	71.1	568	1 TCPQ_YEAST	P47079 saccharomyc
16	32	71.1	679	1 HS82_AJECA	P33125 ajellomyces
17	32	71.1	738	1 STL1_YEAST	P23561 saccharomyc
18	31	68.9	106	1 Y1FL_VIBCH	Q9kkz9 vibrio chol
19	31	68.9	214	1 LIPE_CVACA	O19898 cyanidium c
20	31	68.9	214	1 YBF7_YEAST	P34222 saccharomyc
21	31	68.9	252	1 YGGF_ECOLI	P25894 escherichia
22	31	68.9	260	1 TRPC_LACCA	P17217 lactobacill
23	31	68.9	323	1 HEM2_HELPJ	Q9zmr8 helicobacte
24	31	68.9	323	1 SYK3_PASMU	P57824 pasteurella
25	31	68.9	338	1 SMTA_ASTBI	P56707 astragalus
26	31	68.9	349	1 DMCL_LILLO	P37384 lilium long
27	31	68.9	410	1 VCLB_PEA	P02854 pisum sativ
28	31	68.9	440	1 MTN5_NOSS7	P35678 nostoc sp.
29	31	68.9	486	1 HSL_HUMAN	P14317 homo sapien
30	31	68.9	506	1 YNW3_YEAST	P33867 saccharomyc
31	31	68.9	532	1 COX1_RHOCA	P98059 rhodobacter
32	31	68.9	613	1 ISPG_CHLPN	Q928h0 chlamydia p
33	31	68.9	624	1 YGLO_YEAST	P53137 saccharomyc

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate 1) (HCV).			
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OC	Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -.			
DR	PIR; A39166; GNWVC3.			
DR	HSSP; P27958; IHEI.			
DR	MEROPS; S29.001; -.			
DR	MEROPS; U39.001; -.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002531; HCV_NS1.			
DR	InterPro; IPR002518; HCV_NS2.			
DR	InterPro; IPR004109; HCV_NS3.			
DR	InterPro; IPR000745; HCV_NS4a.			

P46598 candida alb
O13764 schizosacch
P55201 homo sapien
P40220 gallus galli
P29762 homo sapien
P02695 mus musculu
P31709 erwania car
P30041 h antioxi
O08709 mus musculu
O35244 r antioxi
P05325 methanococc
O28058 archaeoglob

DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS4a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01538; HCV_NS1; 1.
 DR Pfam: PF01537; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RDRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00492; HELIC_C3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.94;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVAGLGINAV 10
 Db 1406 KLVAGLGINAV 1415
 RESULT 2

POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain."
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding."
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by a third party for commercial
 CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 CC EMBL: M67463; AAA45534.1; -
 DR PIR: A36814; GNMVCH.
 DR PDB: 1HEI; 25-NOV-98.
 DR PDB: 1A1V; 16-FEB-99.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 746
FT CHAIN 747 809
FT CHAIN 810 1026
FT CHAIN 1027 1657
FT CHAIN 1658 1711
FT CHAIN 1712 1972
FT CHAIN 1973 2420
FT CHAIN 2421 3011
FT TRANSHEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVALGINAV 10
Db 1406 KLVALGINAV 1415

RESULT 3
HS90_PODAN
ID HS90_PODAN STANDARD; PRT; 701 AA.
AC Q43109;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Heat shock protein 90 homolog (Suppressor of vegetative
DE incompatibility MOD-E).
GN MOD-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriales; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S;
RX MEDLINE=97476309; PubMed=9335595;
RA Loubradou G., Begueret J., Turcq B.;
RT "A mutation in an HSP90 gene affects the sexual cycle and suppresses
RT vegetative incompatibility in the fungus Podospora anserina.";
RL Genetics 147:581-588(1997).
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01165; AAB97626.1; -.
DR HSP; P07900; IYER.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR001404; HSP90.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00183; HSP90; 1.
DR PRINTS; PR00775; HEATSHOCK90.
DR SMART; SM00387; HATPase_c; 1.
DR PROSITE; PS00298; HSP90; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 701 AA; 79322 MW; E98A6A541406AACD CRC64;
Query Match 77.8%; Score 35; DB 1; Length 701;
Best Local Similarity 87.5%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLVALGIN 8
Db 662 KLVALGLN 669
RESULT 4
POLG_HCVBK
ID POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein c (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 RN NCBI_TaxID=11105;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE=91140698; PubMed=1847440;
 RA Takizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RA "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RN SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RA "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moonaw E.W., Adachi T., Hostomsky Z.;
 RA "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RA "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE ENVELOPE OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M58335; AAA72945.1; -;
 DR PIR: A38465; GNWTC.
 DR PDB: 1A1Q; 25-MAR-98.
 DR PDB: 1JXP; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 PFam: PF01543; HCV_capsid; 1.

DR PFam: PF01542; HCV_core; 1.
 DR PFam: PF01539; HCV_env; 1.
 DR PFam: PF01560; HCV_NS1; 1.
 DR PFam: PF01538; HCV_NS2; 1.
 DR PFam: PF02907; HCV_NS3; 1.
 DR PFam: PF01006; HCV_NS4a; 1.
 DR PFam: PF01001; HCV_NS4b; 1.
 DR PFam: PF01506; HCV_NS5a; 1.
 DR PFam: PF00998; HCV_RDRP; 1.
 DR PFam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;

Query Match 77.8%; Score 35; DB 1; Length 3010;
 Best Local Similarity 80.0%; Pred. No. 91;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KLVALGINAV 10
 || |||||
 DB 1406 KLSGLGINAV 1415

RESULT 5
 POLG_HCVJA
 ID POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (BC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (BC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088530; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K., Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome."; FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: D90208; BAA14233.1; -.
CC PIR: A39253; GNWVCJ.
CC HSSP: P26663; LUXP.
CC MEROPS: U39.001; -.
CC
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRp.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; HCV_RdRp; 1.
CC Pfam: PF00271; Helicase_C; 1.
CC ProDom: PD186062; HCV_NS1; 1.

DR SMART; SM00492; HELIC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease. INIT_MET 1 1
FT 1 115
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
Query Match 77.8%; Score 35; DB 1; Length 3010;
Best Local Similarity 70.0%; Pred. No. 91;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 KLVALGINAV 10
D 1406 KLTGLGLNAV 1415
RESULT 6
POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,

RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position. Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D11168; BAA01943.1; -;
 DR PIR; A45573; A45573.
 DR HSSP; P26663; 1JXP.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RDRP; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00492; HELICC3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2012
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1155 1165

FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196
 FT CARBOHYD 209 196
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 9A1C77435D642BB CRC64;
 Query Match 77.8%; Score 35; DB 1; Length 3010;
 Best Local Similarity 80.0%; Pred. No. 91;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KLVAIGINAV 10
 Db 1406 KLSGLGINAV 1415
 RESULT 7
 ID DMCL_ARATH STANDARD; PRT; 342 AA.
 AC Q39009;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Meiotic recombination protein DMCL homolog.
 GN LIM15.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96093929; PubMed=7584052;
 RA Sato S., Hotta Y., Tabata S.;
 RT "Structural analysis of a recA-like gene in the genome of Arabidopsis
 RT thaliana.";
 RL DNA Res. 2:89-93(1995).
 CC -!- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- TISSUE SPECIFICITY: TESTIS.
 CC -!- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 CC PROKARYOTIC RECA PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D45415; BAA08255.1; -;
 DR HSSP; P03017; 2REB.
 DR InterPro; IPR001553; RecA.
 DR PROSITE; PS50162; RECA_2; 1.

DR PROSITE; PS50163; RECA_3; 1.
 KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
 FT NP_BIND 133 140 ATP (POTENTIAL).
 SQ SEQUENCE 342 AA; 37339 MW; 8AF0D3D7F7C6750B CRC64;

Query Match 75.6%; Score 34; DB 1; Length 342;
 Best Local Similarity 77.8%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLVALGINA 9
 ||:| ||||
 Db 33 KLIAQGGINA 41

RESULT 8
 DMC1_SOYBN STANDARD; PRT; 345 AA.
 AC Q96449;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Meiotic recombination protein DMC1 homolog.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hadi M.Z., Finer J.J.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO PROKARYOTIC RECA PROTEIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U66836; AAB07025.1; -
 DR HSSP; P03017; 2REB.
 DR InterPro; IPR001553; RECA.
 DR PROSITE; PS50162; RECA_2; 1.
 DR - PROSITE; PS50163; RECA_3; 1.
 KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
 FT NP_BIND 134 141 ATP (POTENTIAL).
 SQ SEQUENCE 345 AA; 37475 MW; 685CB696F923BAB6 CRC64;

Query Match 75.6%; Score 34; DB 1; Length 345;
 Best Local Similarity 77.8%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLVALGINA 9
 ||:| ||||
 Db 34 KLIAQGGINA 42

RESULT 9
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [contains: capsid protein c (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M84754; -; NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNMVTV.
 DR HSSP; P26663; LJXP.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RDRP; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1307 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 75.6%; Score 34; DB 1; Length 3010;

Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KLVALGINAV 10

Db 1406 KLSALGINAV 1415

RESULT 10

ID MMUM_ECOLI STANDARD; PRT; 310 AA.
 AC Q47690; P77226;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Homocysteine S-methyltransferase (EC 2.1.1.10) (S-methylmethionine:homocysteine methyltransferase).
 GN MMUM OR B0261.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,

RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
 RA Mizuno T., Makino K., Nakata A., Yura T., Sampaio G., Mizobuchi K.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the
 RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
 RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99102233; PubMed=9882684;
 RT Thanbichler M., Neuhierl B., Boeck A.;
 RT "S-methylmethionine metabolism in Escherichia coli.";
 RL J. Bacteriol. 181:662-665(1999).
 RN [5]
 RP CHARACTERIZATION.
 RC STRAIN=K12 / JM109;
 RX MEDLINE=99150319; PubMed=10026151;
 RT Neuhierl B., Thanbichler M., Lottspeich F., Boeck A.;
 RT "A family of S-methylmethionine-dependent thiol/selenol
 RT methyltransferases. Role in selenium tolerance and evolutionary
 RT relation.";
 RL J. Biol. Chem. 274:5407-5414(1999).
 CC -!- FUNCTION: CATALYZES METHYL TRANSFER FROM S-METHYLMETHIONINE OR S-
 CC ADENOSYLMETHIONINE (LESS EFFICIENT) TO HOMOCYSTEINE.
 CC SELENOHOMOCYSTEINE AND LESS EFFICIENTLY SELENOCYSTEINE.
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + L-homocysteine = S-
 CC adenosyl-L-homocysteine + L-methionine.
 CC -!- COFACTOR: ZINC (POTENTIAL).
 CC -!- SUBUNIT: MONOMER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000134; AAC73364.1; -;
 DR EMBL; D83536; BAA77929.1; ALT_INIT.
 DR EMBL; U70214; BAB08682.1; -;
 DR EcoGene; EG13343; mmum.
 DR InterPro; IPR003726; S_methyl_trans.
 DR Pfam; PF02574; S_methyl_trans; 1.
 KW Methionine biosynthesis; Transferase; Methyltransferase; Zinc;
 KW Complete proteome.
 FT METAL 229 229 ZINC (POTENTIAL).
 FT METAL 295 295 ZINC (POTENTIAL).
 FT METAL 296 296 ZINC (POTENTIAL).
 FT METAL 123 123 G -> R (IN REF. 2).
 FT CONFLICT 130 130 A -> T (IN REF. 2).
 FT CONFLICT 142 142 S -> T (IN REF. 2).
 FT CONFLICT 159 159 D -> V (IN REF. 2).
 SQ SEQUENCE 310 AA; 33422 MW; 8381CFE475E5FB7A CRC64;

Query Match 73.3%; Score 33; DB 1; Length 310;

Best Local Similarity 60.0%; Pred. No. 27;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KLVALGINAV 10

Db 221 QVVALGINCI 230

RESULT 11

INA_FELCA STANDARD; PRT; 194 AA.
 ID INA_FELCA

P35849; Q28831;
 01-JUN-1994 (Rel. 29, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 Interferon alpha precursor (IFN-alpha).
 Felis silvestris catus (Cat).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 NCBI_TaxID=9685;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=92323151; PubMed=1377975;
 Nakamura N., Sudo T., Matsuda S., Yanai A.;
 "Molecular cloning of feline interferon cDNA by direct expression.";
 Biosci. Biotechnol. Biochem. 56:211-214(1992).
 [2]
 SEQUENCE OF 1-193 FROM N.A.
 MEDLINE=93291263; PubMed=7685640;
 Ueda Y., Sakurai T., Yanai A.;
 "Homogeneous production of feline interferon in silkworm by replacing
 single amino acid code in signal peptide region in recombinant
 baculovirus and characterization of the product.";
 J. Vet. Med. Sci. 55:251-258(1993).
 CC -1- FUNCTION: PRODUCED BY MACROPHAGES, IFN-ALPHA HAVE ANTIVIRAL
 ACTIVITIES. INTERFERON STIMULATES THE PRODUCTION OF TWO ENZYMES:
 A PROTEIN KINASE AND AN OLIGOADENYLATE SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S62636; AAB27160.2; -;
 PIR; JS0664; JS0664.
 DR HSP; P01563; 2HIE.
 DR InterPro; IPR000471; Interferon_abd.
 DR Pfam; PF00143; Interferon; 1.
 DR PRINTS; PR00266; INTERFERONAB.
 DR ProDom; PD000550; Interferon_abd; 1.
 DR SMART; SM00076; IFabd; 1.
 DR PROSITE; PS00252; INTERFERON_A_B_D; 1.
 KW Cytokine; Antiviral; Signal.
 FT SIGNAL 1 BY SIMILARITY.
 FT CHAIN 24 194 INTERFERON ALPHA.
 FT DISULFID 24 123 BY SIMILARITY.
 FT DISULFID 52 166 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 21 21 S -> V (IN REF. 2).
 SQ SEQUENCE 194 AA; 21892 MW; D10E916E3755BFEF CRC64;
 Query Match 71.1%; Score 32; DB 1; Length 194;
 Best Local Similarity 77.8%; Pred. No. 28;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LVALGINAV 10
 Db 11 LVALGCVNSV 19
 RESULT 12
 HS82_ASPFU STANDARD; PRT; 441 AA.
 AC P40292;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Heat shock protein HSP1 (65 kDa IgE-binding protein) (Allergen Asp f

DE 12) (Fragment).
 GN HSP1.
 OS Aspergillus fumigatus (Sartorya fumigata).
 GN Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5085;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Kurup V.P., Banerjee B.;
 RP Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE OF 165-284 FROM N.A.
 RC STRAIN=ATCC 42202 / AF-102;
 RX MEDLINE=93260199; PubMed=8491935;
 RA Kumar A., Reddy L.V., Sochanik A., Kurup V.P.;
 RT "Isolation and characterization of a recombinant heat shock protein
 of Aspergillus fumigatus.";
 RL J. Allergy Clin. Immunol. 91:1024-1030(1993).
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U92465; AAB51544.1; -;
 DR InterPro; IPR001404; HSP90.
 DR Pfam; PF00183; HSP90; 1.
 DR PROSITE; PS00298; HSP90; PARTIAL.
 KW Chaperone; ATP-binding; Heat shock; Allergen.
 FT NON_TER 1
 SQ SEQUENCE 441 AA; 50523 MW; 3726AFF5CCBF83A4 CRC64;
 Query Match 71.1%; Score 32; DB 1; Length 441;
 Best Local Similarity 75.0%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLVALGIN 8
 Db 402 KLVSIGLN 409
 RESULT 13
 GLU4_ORYSA STANDARD; PRT; 499 AA.
 ID GLU4_ORYSA
 AC P14323;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Glutelin precursor.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. NATO CT8998;
 RX MEDLINE=90067963; PubMed=2587279;
 RA Wen L., Huang J.K., Johnson B.H., Reek G.R.;
 RT "Nucleotide sequence of a cDNA that encodes a rice glutelin.";
 RL Nucleic Acids Res. 17:9490-9490(1989).
 CC -1- FUNCTION: THIS IS A SEED STORAGE PROTEIN.
 CC -1- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
 BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
 DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEIN (GLOBULINS)

CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15833; CAA33838.1; -
CC PIR; S07640; S07640.
CC InterPro; IPR000459; Seedstore_11s.
CC Pfam; PF00190; Seedstore_11s; 1.
CC PRINTS; PR00439; 11SGLOBULIN.
CC PROSITE; PS00305; 11S_SEED_STORAGE; 1.
CC Seed storage protein; Signal; Multigene family.
CC SIGNAL 1 24
CC CHAIN 25 302 ACIDIC CHAIN.
CC CHAIN 303 499 BASIC CHAIN.
CC DISULFID 121 309 INTERCHAIN (ACIDIC-BASIC) (POTENTIAL).
CC SEQUENCE 499 AA; 56564 MW; 38EA5A7398EEAA15 CRC64;
CC -----
CC Query Match 71.1%; Score 32; DB 1; Length 499;
CC Best Local Similarity 100.0%; Pred. No. 67;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 4 ALGINAV 10
CC DB 238 ALGINAV 244
CC -----
CC RESULT 14
CC GLUB_ORYSA
CC ID TCPO_YEAST STANDARD; PRT; 499 AA.
CC AC Q02898;
CC DT 01-JUN-1994 (Rel. 29, Created)
CC DT 01-JUN-1994 (Rel. 29, Last sequence update)
CC DT 01-JUN-1994 (Rel. 29, Last annotation update)
CC DE Glutelin type-B 1 precursor.
CC GN GLUB-1.
CC OS Oryza sativa (Rice).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
CC OC Eriarthoideae; Oryzeae; Oryza.
CC OX NCBI_TaxID=4530;
CC [1]
CC SEQUENCE FROM N.A.
CC RC STRAIN=CV. MANGETSUMOCHI; TISSUE=Seed;
CC RX MEDLINE=92003697; PubMed=1680490;
CC RA Takaiwa F., Oono K., Wing D., Kato A.;
CC RT "Sequence of three members and expression of a new major subfamily of
CC RT glutelin genes from rice."
CC RL Plant Mol. Biol. 17:875-885(1991).
CC CC -!- FUNCTION: THIS IS A SEED STORAGE PROTEIN.
CC CC -!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC CC DISULFIDE BOND.
CC CC -!- DEVELOPMENTAL STAGE: IT BEGINS TO ACCUMULATE 6 DAYS AFTER
CC CC FLOWERING AND REACHES A MAXIMUM LEVEL AT 14 DAYS.
CC CC -!- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEIN (GLOBULINS)
CC CC FAMILY. GLUTELIN GENE SUBFAMILY B.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54314; CAA38212.1; -
CC

DR PIR; S17762; S17762.
DR InterPro; IPR000459; Seedstore_11s.
DR Pfam; PF00190; Seedstore_11s; 1.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
KW Seed storage protein; Signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 25 302 ACIDIC CHAIN (BY SIMILARITY).
FT CHAIN 303 499 BASIC CHAIN (BY SIMILARITY).
FT DISULFID 121 309 INTERCHAIN (ACIDIC-BASIC) (POTENTIAL).
FT SEQUENCE 499 AA; 56675 MW; 67582414D75166AB CRC64;
FT -----
FT Query Match 71.1%; Score 32; DB 1; Length 499;
FT Best Local Similarity 100.0%; Pred. No. 67;
FT Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT -----
FT QY 4 ALGINAV 10
FT DB 238 ALGINAV 244
FT -----
FT RESULT 15
FT TCPO_YEAST
FT ID TCPO_YEAST STANDARD; PRT; 568 AA.
FT AC P47079;
FT DT 01-NOV-1995 (Rel. 32, Created)
FT DT 01-NOV-1995 (Rel. 32, Last sequence update)
FT DT 01-OCT-1996 (Rel. 34, Last annotation update)
FT DE T-complex protein 1, theta subunit (TCP-1-theta) (CCT-theta).
FT GN CCT8 OR YJL008C OR J1374.
FT OS Saccharomyces cerevisiae (Baker's yeast).
FT OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
FT OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
FT OX NCBI_TaxID=4932;
FT [1]
FT SEQUENCE FROM N.A.
FT RA To Van D., Perea J., Jacq C.;
FT RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
FT CC -!- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
FT CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
FT CC ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
FT CC FORMATION (BY SIMILARITY).
FT CC -!- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 kDa THAT
FT CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER (BY SIMILARITY).
FT CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
FT CC -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
FT CC -----
FT This SWISS-PROT entry is copyright. It is produced through a collaboration
FT between the Swiss Institute of Bioinformatics and the EMBL outstation -
FT the European Bioinformatics Institute. There are no restrictions on its
FT use by non-profit institutions as long as its content is in no way
FT modified and this statement is not removed. Usage by and for commercial
FT entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
FT or send an email to license@isb-sib.ch).
FT -----
FT EMBL; Z49284; CAA89300.1; -
FT HSSP; P48424; 1A6D.
FT SGD; S0003545; CCT8.
FT InterPro; IPR002423; TCP1_cpn60.
FT InterPro; IPR002194; TCP1_1.
FT Pfam; PF00118; cpn60_TCP1; 1.
FT PRINTS; PR00304; TCOMPLEXTCP1.
FT PROSITE; PS00750; TCP1_1; 1.
FT PROSITE; PS00751; TCP1_2; 1.
FT PROSITE; PS00995; TCP1_3; 1.
KW Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 568 AA; 61662 MW; 90817CA3151A52FD CRC64;
SQ -----
SQ Query Match 71.1%; Score 32; DB 1; Length 568;
SQ Best Local Similarity 50.0%; Pred. No. 76;
SQ Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVALGINAV 10
| | : : | : | |
Db 116 KLISMGLSAV 125

Search completed: August 23, 2002, 09:55:24
Job time: 54 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:21:59 ; Search time 139.83 seconds
(without alignments)
12.372 Million cell updates/sec

Title: US-08-854-825-28

Perfect score: 45

Sequence: 1 KLVALGINAV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	45	100.0	253	12	055407 hepatitis c
2	45	100.0	253	12	055408 hepatitis c
3	45	100.0	253	12	055409 hepatitis c
4	45	100.0	253	12	055410 hepatitis c
5	45	100.0	253	12	055411 hepatitis c
6	45	100.0	253	12	055412 hepatitis c
7	45	100.0	253	12	055413 hepatitis c
8	45	100.0	253	12	055414 hepatitis c
9	45	100.0	253	12	055416 hepatitis c
10	45	100.0	253	12	055417 hepatitis c
11	45	100.0	253	12	055418 hepatitis c
12	45	100.0	253	12	055419 hepatitis c
13	45	100.0	253	12	055420 hepatitis c
14	45	100.0	253	12	055422 hepatitis c
15	45	100.0	253	12	055423 hepatitis c
16	45	100.0	253	12	055424 hepatitis c
					055425 hepatitis c

17	45	100.0	253	12	055426	055426 hepatitis c
18	45	100.0	253	12	055427	055427 hepatitis c
19	45	100.0	253	12	055428	055428 hepatitis c
20	45	100.0	253	12	055429	055429 hepatitis c
21	45	100.0	253	12	055430	055430 hepatitis c
22	45	100.0	253	12	055431	055431 hepatitis c
23	45	100.0	253	12	055432	055432 hepatitis c
24	45	100.0	253	12	055433	055433 hepatitis c
25	45	100.0	253	12	055434	055434 hepatitis c
26	45	100.0	253	12	055435	055435 hepatitis c
27	45	100.0	253	12	055436	055436 hepatitis c
28	45	100.0	253	12	055437	055437 hepatitis c
29	45	100.0	253	12	055438	055438 hepatitis c
30	45	100.0	253	12	055439	055439 hepatitis c
31	45	100.0	253	12	055440	055440 hepatitis c
32	45	100.0	253	12	055441	055441 hepatitis c
33	45	100.0	253	12	055442	055442 hepatitis c
34	45	100.0	253	12	055443	055443 hepatitis c
35	45	100.0	253	12	055444	055444 hepatitis c
36	45	100.0	253	12	055445	055445 hepatitis c
37	45	100.0	253	12	055446	055446 hepatitis c
38	45	100.0	253	12	055447	055447 hepatitis c
39	45	100.0	253	12	055448	055448 hepatitis c
40	45	100.0	253	12	055449	055449 hepatitis c
41	45	100.0	253	12	055450	055450 hepatitis c
42	45	100.0	253	12	055451	055451 hepatitis c
43	45	100.0	253	12	055452	055452 hepatitis c
44	45	100.0	253	12	055453	055453 hepatitis c
45	45	100.0	253	12	055454	055454 hepatitis c

ALIGNMENTS

RESULT 1

OS5407	PRELIMINARY;	PRT;	253 AA.
ID	OS5407		
AC	OS5407;		
DT	01-JUN-1998 (TrEMBLrel. 06, Created)		
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	POLYPROTEIN (FRAGMENT).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wang H., Eckels D.D.;		
RT	*Sequence analysis of HCV NS3 variants from a patient infected with		
RT	HCV.*;		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF035122; AAB88140.1; -		
DR	HSSP; P27958; IHEI.		
DR	MEROPS; S29.001; -		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR001650; Helicase_C.		
DR	Pfam; PF00271; Helicase_C; 1.		
KW	ATP-binding; Helicase.		
FT	NON_TER 1		
FT	NON_TER 253		
SQ	SEQUENCE 253 AA; 26732 MW; 780BC94A2A9DF1B2 CRC64;		

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10

Db 109 KLVALGINAV 118

```

RESULT 2
O55408
ID O55408 PRELIMINARY; PRT; 253 AA.
AC O55408;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035123; AAB88141.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00490; HELICc; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26895 MW; CFB748C19A98DBC1 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 3
O55409
ID O55409 PRELIMINARY; PRT; 253 AA.
AC O55409;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035124; AAB88142.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26839 MW; 4E2C980D5598D8D1 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 4
O55410
ID O55410 PRELIMINARY; PRT; 253 AA.
AC O55410;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035125; AAB88143.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26828 MW; 4F882749AB72997B CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 5
O55411
ID O55411 PRELIMINARY; PRT; 253 AA.
AC O55411;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035126; AAB88144.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26871 MW; 994B15908EB8D578 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

```

```
Query Match      100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 6
O55413 PRELIMINARY; PRT; 253 AA.
AC O55413;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035128; AAB88146.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26851 MW; 4E2913163990C8D1 CRC64;

Query Match      100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 7
O55414 PRELIMINARY; PRT; 253 AA.
AC O55414;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035129; AAB88147.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.

Query Match      100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 8
O55416 PRELIMINARY; PRT; 253 AA.
AC O55416;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035131; AAB88149.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26811 MW; 4E2C980D558A53D1 CRC64;

Query Match      100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118

RESULT 9
O55417 PRELIMINARY; PRT; 253 AA.
AC O55417;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26831 MW; 74389E3256BAFD98 CRC64;

Query Match      100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVALGINAV 10
Db 109 KLVALGINAV 118
```

DR EMBL; AF035132; AAB88150.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR ATP-binding; Helicase.
KW ATP-binding; Helicase.
FT NON_TER 253 253
SQ SEQUENCE 253 AA; 26839 MW; 4E2C980D5598D8D1 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 109 KLVALGINAV 118

RESULT 10
O55418 PRELIMINARY; PRT; 253 AA.
AC O55418;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035133; AAB88151.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 253 253
SQ SEQUENCE 253 AA; 26849 MW; 1FFE49DE559BD99A CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 109 KLVALGINAV 118

RESULT 11
O55419 PRELIMINARY; PRT; 253 AA.
AC O55419;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]

RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035134; AAB88152.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 253 253
SQ SEQUENCE 253 AA; 26840 MW; 21143ED26FDB7CEB CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 109 KLVALGINAV 118

RESULT 12
O55420 PRELIMINARY; PRT; 253 AA.
AC O55420;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with HCV."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035135; AAB88153.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 253 253
SQ SEQUENCE 253 AA; 26897 MW; CC00BACF797925FD CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVALGINAV 10
|||||

Db 109 KLVALGINAV 118

RESULT 13
O55422 PRELIMINARY; PRT; 253 AA.
AC O55422;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).

OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035137; AAB88155.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26867 MW; 12C186C261F8964E CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVALGINAV 10
DB 109 KLVALGINAV 118
|||||

RESULT 14
O55423 PRELIMINARY; PRT; 253 AA.
AC O55423;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035138; AAB88156.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26892 MW; 1FE43BB6EBE13F4F CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVALGINAV 10
DB 109 KLVALGINAV 118
|||||

RESULT 15
O55424 PRELIMINARY; PRT; 253 AA.
ID O55424

O55424;
AC 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Eckels D.D.;
RT "Sequence analysis of HCV NS3 variants from a patient infected with
RT HCV";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035139; AAB88157.1; -.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
FT NON_TER 253
SQ SEQUENCE 253 AA; 26882 MW; 0A9693B336EAA303 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVALGINAV 10
DB 109 KLVALGINAV 118
|||||

Search completed: August 23, 2002, 10:21:59
Job time: 1499 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:01:01 ; Search time 179.72 Seconds
(without alignments)
5.562 Million cell updates/sec

Title: US-08-854-825-42
Perfect score: 45
Sequence: 1 ILDSFDPLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	9	16	AA84577
2	45	100.0	9	20	AAV10237
3	45	100.0	9	20	AAV10518
4	45	100.0	44	13	AA120755
5	45	100.0	47	13	AA120763
6	45	100.0	55	13	AA120762
7	45	100.0	55	18	AAW01867
8	45	100.0	126	10	AA190151
9	45	100.0	126	10	AA192034
10	45	100.0	195	15	AA192660
11	45	100.0	411	17	AA190934

12	45	100.0	499	19	AAW67010	HCV non-structural
13	45	100.0	504	12	AA14553	Chimpanzee HCV clo
14	45	100.0	504	15	AA62659	Hepatitis C virus
15	45	100.0	504	16	AA81941	Hepatitis C virus
16	45	100.0	504	20	AAW70825	Hepatitis C virus
17	45	100.0	504	20	AAW80513	Hepatitis C virus
18	45	100.0	541	13	AA121567	HCV CKS-NS5F - pHc
19	45	100.0	541	14	AA133634	HCV CKS-NS5F fusio
20	45	100.0	541	14	AA133576	HCV CKS-NS5F recom
21	45	100.0	541	14	AA133596	HCV CKS-NS5F recom
22	45	100.0	541	22	AA151374	HCV recombinant an
23	45	100.0	798	14	AA133630	HCV CKS-NS5 EF fus
24	45	100.0	798	22	AA169012	HCV recombinant an
25	45	100.0	1766	10	AA192041	Sequence encoded i
26	45	100.0	1771	22	AA162631	HCV NS35 polypepti
27	45	100.0	1771	22	AA162632	HCV delNS35 polype
28	45	100.0	1771	22	AA162634	Amino acid sequenc
29	45	100.0	1771	22	AA162635	Amino acid sequenc
30	45	100.0	1786	10	AA190158	Protein sequence o
31	45	100.0	1892	22	AA162636	Amino acid sequenc
32	45	100.0	1911	22	AA162638	Amino acid sequenc
33	45	100.0	1921	22	AA162639	Amino acid sequenc
34	45	100.0	1944	22	AA162637	Amino acid sequenc
35	45	100.0	2261	10	AA190164	Peptide encoded by
36	45	100.0	2301	10	AA192047	Sequence encoded i
37	45	100.0	2435	13	AA125135	HCV polypeptide 1.
38	45	100.0	2436	10	AA192050	Sequence encoded i
39	45	100.0	2436	10	AA190288	Peptide encoded by
40	45	100.0	2436	13	AA128582	HCV amino acid seq
41	45	100.0	2772	11	AA108123	Hepatitis C virus
42	45	100.0	2772	21	AA181540	Protein encoded by
43	45	100.0	2816	14	AA134009	HCV-1 polyprotein.
44	45	100.0	2894	13	AA124440	Composite HCV HC-J
45	45	100.0	2894	16	AA170230	Composite hepatitis

ALIGNMENTS

RESULT 1
ID AAR84577 standard; peptide; 9 AA.
XX AAR84577;
DT 25-APR-1996 (first entry)
XX Cytotoxic T-cell epitope, aa 2252-2260 of HCV-1 NS5 region.
DE Hepatitis C virus; HCV; epitope; vaccine; immunogen.
XX Hepatitis C virus.
OS Hepatitis C virus.
XX WO9525122-A1.
XX 21-SEP-1995.
XX 16-MAR-1995; 95WO-US03224.
XX 17-MAR-1994; 94US-0214650.
XX (SCRI) SCRIPPS RES INST.
XX Cerny A, Chisari EV;
XX WPI; 1995-336941/43.
XX Novel molecule comprising a cytotoxic T cell epitope - used to
PT vaccinate against hepatitis C viral infection
XX Claim 1; Page 71; 85pp; English.
XX AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides

CC from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
 CC were tested for peptide specific cytotoxic T-cell activity. The
 CC peptides AAR84570-77 were found to have substantial homology with a T-
 CC cell epitope and are useful in vaccines against HCV infection.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 16; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 1 Ildsfplv 9
 |||||

RESULT 2
 AAY10237
 ID AAY10237 standard; Peptide; 9 AA.

XX AC AAY10237;

XX DT 12-MAY-1999 (first entry)

XX DE T cell epitope/MHC ligand SEQ ID NO:167.

XX KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX OS Synthetic.

XX OS Hepatitis C virus.

XX PN WO9902183-A2.

XX XX 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-US14289.

XX PR 10-DEC-1997; 97US-0988320.

XX PR 10-JUL-1997; 97CA-2209815.

XX PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX PI Kuendig TM, Simard JUL;

XX XX WPI; 1999-120514/10.

XX Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX PS Disclosure; Page 30; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 1 Ildsfplv 9
 |||||

RESULT 3
 AAY10518
 ID AAY10518 standard; Peptide; 9 AA.

XX AC AAY10518;

XX DT 12-MAY-1999 (first entry)

XX DE HLA Class I motif peptide SEQ ID NO:448.

XX KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 KW immunisation; tumour; infectious disease; immunotherapy; cancer;
 KW malignant melanoma; viral disease; hepatitis; AIDS.

XX OS Synthetic.

XX OS Hepatitis C virus.

XX PN WO9902183-A2.

XX PD 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-US14289.

XX PR 10-DEC-1997; 97US-0988320.

XX PR 10-JUL-1997; 97CA-2209815.

XX PA (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX PI Kuendig TM, Simard JUL;

XX XX WPI; 1999-120514/10.

XX Inducing a cytotoxic T lymphocyte response - by maintaining a level
 PT of antigen in the lymphatic system of a mammal so as to provide a
 PT sustained CTL response, used to treat, e.g. AIDS

XX PS Disclosure; Page 44; 199pp; English.

XX The present invention describes a method of inducing and/or sustaining
 CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 CC method comprises: (a) delivering an antigen to the mammal at a level to
 CC induce an immunological CTL response in the mammal; and (b) maintaining
 CC the level of the antigen in the mammal's lymphatic system to maintain
 CC the immunologic CTL response. The method can be used for the delivery of
 CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 CC gene antigen, or a viral antigen. They can be used for the treatment of
 CC disease such as cancer, e.g. malignant melanoma or infectious disease,
 CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 CC to the lymphatic system provides for potent CTL stimulation that takes
 CC place in the milieu of the lymphoid organ, and it sustains stimulation
 CC that is necessary to keep CTL active, cytotoxic and recirculating
 CC through the body. AAY10071 to AAY10639 represent examples of peptide
 CC antigens given in the present invention.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILDSFDPLV 9
 |||||
Db 1 ildsfdplv 9

RESULT 4

AAR20755
ID AAR20755 standard; Protein; 44 AA.

XX AC AAR20755;
XX DT 05-MAY-1992 (first entry)

XX DE Peptide 5 based on immunoreactive region of Hepatitis C virus.

XX KW Non-A, non-B hepatitis virus; non-structural protein; vaccine.
XX OS Synthetic.

XX PN EP468527-A.

XX PD 29-JAN-1992.

XX PF 26-JUL-1991; 91EP-0112620.

XX PR 24-JUN-1991; 91US-0719819.

XX PR 26-JUL-1990; 90US-0558799.

XX PR 07-FEB-1991; 91US-0651735.

XX PR 11-MAR-1991; 91US-0667275.

XX PA (UNBI-) UTD BIOMEDICAL INC.

XX PI Chang YW, Hosein B;

XX DR WPI; 1992-034279/05.

XX PT New synthetic peptide specific for HCV antibodies - for detection
XX PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
XX PT immunogen for preparation of vaccines

XX PS Claim 1; Page 88; 98pp; English.

XX CC This peptide is one of 19 specifically claimed antigens based on the
XX CC immunoreactive regions of the envelope protein and non-structural
XX CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
XX CC invention also concerns analogues, segments, mixtures, conjugates
XX CC and polymers of these peptides. The C-terminal amino acid may be
XX CC amidated. See AAR20751-R20782.

XX SQ Sequence 44 AA;

Query Match 100.0%; Score 45; DB 13; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.057;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILDSFDPLV 9
 |||||
Db 30 ildsfdplv 38

RESULT 5

AAR20763
ID AAR20763 standard; Protein; 47 AA.

XX AC AAR20763;

XX DT 05-MAY-1992 (first entry)

XX DE Peptide 13 based on immunoreactive region of Hepatitis C virus.

XX KW Non-A, non-B hepatitis virus; non-structural protein; vaccine.

XX XX

OS Synthetic.

XX PN EP468527-A.

XX PD 29-JAN-1992.

XX PF 26-JUL-1991; 91EP-0112620.

XX PR 24-JUN-1991; 91US-0719819.

XX PR 26-JUL-1990; 90US-0558799.

XX PR 07-FEB-1991; 91US-0651735.

XX PR 11-MAR-1991; 91US-0667275.

XX PA (UNBI-) UTD BIOMEDICAL INC.

XX PI Chang YW, Hosein B;

XX DR WPI; 1992-034279/05.

XX PT New synthetic peptide specific for HCV antibodies - for detection
XX PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
XX PT immunogen for preparation of vaccines

XX PS Claim 1; Page 90; 98pp; English.

XX CC This peptide is one of 19 specifically claimed antigens based on the
XX CC immunoreactive regions of the envelope protein and non-structural
XX CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
XX CC invention also concerns analogues, segments, mixtures, conjugates
XX CC and polymers of these peptides. The C-terminal amino acid may be
XX CC amidated. See AAR20751-R20782.

XX SQ Sequence 47 AA;

Query Match 100.0%; Score 45; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.062;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILDSFDPLV 9
 |||||
Db 19 ildsfdplv 27

RESULT 6

AAR20762
ID AAR20762 standard; Protein; 55 AA.

XX AC AAR20762;

XX DT 05-MAY-1992 (first entry)

XX DE Peptide 12 based on immunoreactive region of Hepatitis C virus.

XX KW Non-A, non-B hepatitis virus; non-structural protein; vaccine.

XX OS Synthetic.

XX PN EP468527-A.

XX PD 29-JAN-1992.

XX PF 26-JUL-1991; 91EP-0112620.

XX PR 24-JUN-1991; 91US-0719819.

XX PR 26-JUL-1990; 90US-0558799.

XX PR 07-FEB-1991; 91US-0651735.

XX PR 11-MAR-1991; 91US-0667275.

XX PA (UNBI-) UTD BIOMEDICAL INC.

XX PI Chang YW, Hosein B;

XX XX

DR WPI; 1992-034279/05.

XX New synthetic peptide specific for HCV antibodies - for detection
PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
PT immunogen for preparation of vaccines

XX Claim 1; Page 89; 98pp; English.

XX This peptide is one of 19 specifically claimed antigens based on the
CC immunoreactive regions of the envelope protein and non-structural
CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
CC invention also concerns analogues, segments, mixtures, conjugates
CC and polymers of these peptides. The C-terminal amino acid may be
CC amidated. See AAR20751-K20782.

XX Sequence 55 AA;

Query Match 100.0%; Score 45; DB 13; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
Db 41 lldsfldplv 49

RESULT 7

AAW01867
ID AAW01867 standard; peptide; 55 AA.

AC AAW01867;

DT 07-NOV-1997 (first entry)

DE HCV NS-5 prototype immunoreactive peptide 2.

XX immunoreactive; HCV; Hepatitis C Virus; antibody; screening; blood;
KW diagnose; infection; immunoassay; core protein; NS-4; NS-5.

XX Synthetic.

XX Key Location/Qualifiers
FH Misc-difference 8 /note= "can be substituted with Ornithine"
FT Misc-difference 16 /note= "can be substituted with Norvaline"
FT Misc-difference 24 /note= "can be substituted with Asparagine"
FT Misc-difference 32 /note= "can be substituted with Ornithine"
FT Misc-difference 40 /note= "can be substituted with Norvaline"
FT Misc-difference 48 /note= "can be substituted with Norleucine"

XX DE19540105-C1.

XX 20-FEB-1997.

XX 27-OCT-1995; 95DE-1040105.

XX 19-SEP-1995; 95US-0530550.

XX (UNBI-) UNITED BIOMEDICAL INC.

PI Hosein B, Wang CY;

XX WPI; 1997-120479/12.

XX Compsn. contg. synthetic peptide(s) reactive with hepatitis C
PT related antibodies - used for screening blood and diagnosis of
PT infection

XX PS

Claim 10; Page 31-32; 46pp; German.

XX This peptide is the prototype peptide for design of synthetic peptides
CC immunoreactive with Hepatitis C virus (HCV) antibodies against the NS-5
CC protein. The synthetic peptides containing substitution as indicated in
CC the FT above, can be used in a novel composition which comprises at least
CC one linear or branched peptide (A) of formulae: P1-Y; P2-X; P4X2X;
CC P8X4X2X; or P16X8X4X2X; Y = COOH or CONH2 at the peptide C terminus; X =
CC residues of amino acids (aa), or analogues with 2 amino gps. and one COOH
CC gp., each gp. able to form a peptide bond; P = peptide that is
CC specifically immunoreactive with HCV antibodies. The compositions are
CC used to detect HCV antibodies, e.g. for screening blood products, and to
CC diagnose HCV infection by immunoassay of serum, tissue (extracts) and
CC body fluids, pref. by ELISA, sandwich or passive haemagglutination tests.

XX Sequence 55 AA;

Query Match 100.0%; Score 45; DB 18; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
Db 41 lldsfldplv 49

RESULT 8

AAP90151
ID AAP90151 standard; protein; 126 AA.

AC AAP90151;

DT 01-NOV-1989 (first entry)

XX Sequence of hepatitis C virus cDNA insert in clone 8f.

XX Hepatitis C virus; clone 8f; clone 14c; probe; vaccine.

XX Pan troglodytes.

XX Key Location/Qualifiers
FH Region 1..24

PN GB2212511-A.

PD 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-215054/30.

XX N-PSDB; AAN90320.

XX Hepatitis C virus gene - used for prodn. of polynucleotide.probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.

PS Disclosure; fig 19; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in clone 8f (see AAN90320). The polypeptides
CC are used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.
CC The region shown overlaps with clone 14c.

XX Sequence 126 AA;

\dot{O}_2

```

OS Hepatitis C virus.
XX EP693687-A1.
XX 24-JAN-1996.
XX 03-APR-1991; 91EP-0114016.
XX 04-APR-1990; 90US-0504352.
XX (CHIR ) CHIRON CORP.
XX Choo Q, Houghton M, Kuo G;
XX WPI; 1996-117956/13.
XX Combinations of synthetic Hepatitis C Virus antigens - provide more
XX effective diagnosis of Non-A, Non-B Hepatitis
XX Claim 8; Fig 1(A-Y); 53pp; English.
XX The combination comprises an HCV antigen from the C domain (pref.
XX C22 - AAR90936) and at least one HCV antigen from the NS3 (pref. C33c
XX - AAR90932), NS4 (pref. C100 - AAR90933), S (pref. S2 - AAR90935) or NS5
XX (AAR90934) domain.
XX The antigens may in the form of a fusion protein, a simple physical
XX mixture, or the individual antigens commonly bound to a solid matrix.
XX They are pref. prepd. by recombinant DNA techniques (primers are
XX given in AAT12711-T12716), but can be synthesised or isolated from
XX HCV using affinity chromatography.
XX Sequence 411 AA;

Query Match 100.0%; Score 45; DB 17; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.92; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 ILDSFDPLV 9
DB 199 ildsfdplv 207

RESULT 12
AAW67010
ID AAW67010 standard; protein; 499 AA.
XX
XX AAW67010;
AC
XX 02-MAR-1999 (first entry)
XX HCV non-structural protein NS4.
XX Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
XX non-structural protein; thioamide bond; peptide bond.
XX Hepatitis C virus.
XX JP10226698-A.
XX 25-AUG-1998.
XX 19-FEB-1997; 97JP-0034702.
XX 19-FEB-1997; 97JP-0034702.
XX (KYOW ) KYOWA MEDEX KK.
XX WPI; 1998-515103/44.
XX Determination of antibody in sample - uses peptide analog absorbed
XX or chemically bound on carrier as antigen

```

```

PS Disclosure; Page 5; 13pp; Japanese.
XX This sequence represents the Hepatitis C virus (HCV) non-structural
XX (NS4) protein. The invention relates to peptide analogues derived from
XX HCV proteins, e.g. AAW67417-W67426, which can be used for the
XX determination of anti-HCV antibodies in a sample. Preferably the peptide
XX analogues contain one or more thioamide peptide bonds where at least one
XX oxygen atom of the peptide bond is replaced by sulphur atom. The
XX peptide analogues can be adsorbed or chemically bound to a carrier.
XX Sequence 499 AA;

Query Match 100.0%; Score 45; DB 19; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
DB 184 ildsfdplv 192

RESULT 13
AAR14553
ID AAR14553 standard; Protein; 504 AA.
XX
XX AAR14553;
AC
XX 24-JAN-1992 (first entry)
XX Chimpanzee HCV clone 36.
XX non-A, non-B hepatitis; NANBH; liver disease; ss.
XX Hepatitis C Virus.
XX WO9115516-A.
XX 17-OCT-1991.
XX 05-APR-1991; 91WO-US02370.
XX 09-OCT-1990; 90US-0594854.
XX 06-APR-1990; 90US-0505611.
XX (GENE-) GENELABS INC.
XX Reyes G, Kim JP, Moeckli R, Simonsen CC;
XX WPI; 1991-325174/44.
XX New hepatitis C virus epitope(s) immuno-reactive with HCV
XX infected sera - useful for detection of HCV infections and as HCV
XX vaccine
XX Claim 2; Fig 4; 150pp; English.
XX cDNA libraries were prepared from the serum of one HCV infected
XX chimpanzee and 4 HCV infected humans. The 5 libraries were
XX immunoscreened using HCV positive human and chimp sera. Inserts from
XX HCV-immunopositive clones derived from chimp sera were characterised
XX as exogenous to normal human PBL DNA. Two of these HCV clones were
XX characterised - Clones 36 and 40. The nucleotide sequence of clone
XX 36 corresponds to nucleotides 5010 to 6516 of the HCV sequence. The
XX lambda gt11 clone of clone 36 in E.coli was deposited as ATCC 40901.
XX See AAL4276 (clone 40).
XX Sequence 504 AA;

Query Match 100.0%; Score 45; DB 12; Length 504;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ILDSFDPLV 9
|||||
Db 133 ildsfplv 141

RESULT 14

AAR62659
ID AAR62659 standard; Protein; 504 AA.

XX AC AAR62659;

XX DT 30-JUN-1995 (first entry)

XX DE Hepatitis Type C virus protein fragment encoded by bases 5010-6516.

XX KW Hepatitis virus; HCV; hepatitis; immunoscreening; parentally; PT-NANBH;
KW parenterally transmitted; non-A, non-B hepatitis; chimpanzee; human;
KW PCR; amplification; amplify; primer; antibody; antigen; detection;
KW immunoprophylaxis; vaccine.

XX OS Hepatitis C-type Virus.

XX PN WO9424565-A.

XX PD 27-OCT-1994.

XX PF 15-APR-1994; 94WO-US04174.

XX PR 22-APR-1993; 93US-0052542.

XX PA (GENE-) GENELABS TECHNOLOGIES INC.

XX PI Kim JP, Moeckli R, Reyes GR;

XX DR WPI; 1994-342055/42.

XX DR N-PSDB; AAQ73003.

XX PT New hepatitis C virus poly:peptide(s) - used for detection and
PT prodn. of antibodies for detection or immuno:prophylaxis, or in
PT vaccines

XX PS Claim 5; Page 60-63; 94pp; English.

XX CC The amino acid sequence of a portion of the type-C hepatitis virus (HCV)
CC The corresponding DNA was isolated by the immunoscreening of lambda-gt11
CC cDNA libraries made from the serum from one parenterally transmitted
CC non-A, non-B hepatitis (PT-NANBH) chimpanzee and 4 PT-NANBH human
CC patients. This clone, designated clone no. 36 covered the sequence from
CC bases 5010-6516 of the published HCV sequence and encodes a polypeptide
CC of 504 amino acids. This clone was inserted into the pET3d vector to
CC generate clone pET-36. The pET-36 clone was used to generate 3
CC overlapping subfragments (AAQ73016-8) by PCR amplification using primers
CC (AAQ73010-5). The polypeptides encoded by the DNA fragments can be used
CC to generate antibodies to HCV antigens. The antibodies are useful for
CC the detection of hepatitis type C antigens or for immunoprophylaxis.
CC The polypeptides can also be used as vaccines effective against HCV.

XX SQ Sequence 504 AA;

Query Match 100.0%; Score 45; DB 15; Length 504;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
|||||
Db 133 ildsfplv 141

RESULT 15

AAR81941

ID AAR81941 standard; Protein; 504 AA.

XX AAR81941;
XX AC 28-JUL-1996 (first entry)
XX DT Hepatitis C virus antigen.
XX DE Hepatitis C virus; HCV; antigen; detection; diagnosis; vaccine;
XX KW antibodies; immunoprophylaxis; sera; serum.
XX OS Hepatitis C virus (Clone 303-1-4).

XX PN US5443965-A.

XX PD 22-AUG-1995.

XX PF 06-APR-1990; 90US-0505611.

XX PR 05-APR-1991; 91US-0681703.

XX PR 06-APR-1990; 90US-0505611.

XX PR 10-OCT-1990; 90US-0594854.

XX PA (GENE-) GENELABS INC.

XX PI Kim JP, Moeckli R, Reyes GR;

XX DR WPI; 1995-302120/39.

XX DR N-PSDB; AAQ98204.

XX PT New nucleic acids encoding hepatitis C virus antigens - used to
PT develop prods. for detection of HCV-infected sera and prodn. of
PT vaccines and anti-HCV antibodies.

XX PS Disclosure; Columns 57-60; 71pp; English.

XX CC Hepatitis C virus (HCV) antigens can be used for detecting HCV
CC infected sera and individuals infected with HCV. They can also be
CC used in an anti-HCV vaccine or for the production of anti-HCV
CC antibodies which can be used for passive immunoprophylaxis. The
CC antigens consistently identify more HCV positive serum samples with
CC a high degree of specificity. See AAQ98202-14 and AAR81939-51.

XX SQ Sequence 504 AA;

Query Match 100.0%; Score 45; DB 16; Length 504;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
|||||
Db 133 ildsfplv 141

Search completed: August 23, 2002, 10:01:01
Job time: 391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:53 ; Search time 61.77 Seconds
(without alignments)
3.559 Million cell updates/sec

Title: US-08-854-825-42
Perfect score: 45
Sequence: 1 ILDSFDPLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	9	1	US-08-214-650-42
2	45	100.0	9	2	US-08-146-028-383
3	45	100.0	9	4	US-08-723-425A-383
4	45	100.0	9	4	US-09-112-206-383
5	45	100.0	35	1	US-08-262-037-103
6	45	100.0	35	1	US-08-262-037-106
7	45	100.0	44	1	US-08-262-037-5
8	45	100.0	47	1	US-08-262-037-13
9	45	100.0	50	1	US-08-262-037-104
10	45	100.0	55	1	US-08-530-550-36
11	45	100.0	55	1	US-08-262-037-12
12	45	100.0	126	4	US-08-444-818-40
13	45	100.0	195	5	PCT-US94-04174-17
14	45	100.0	504	1	US-07-853-985A-6
15	45	100.0	504	1	US-07-681-703B-6
16	45	100.0	504	1	US-08-184-236-6
17	45	100.0	504	2	US-08-407-410B-6
18	45	100.0	504	2	US-08-485-500-6
19	45	100.0	504	5	PCT-US91-02370-6
20	45	100.0	504	5	PCT-US94-04174-6
21	45	100.0	541	1	US-08-867-611-8
22	45	100.0	541	5	PCT-US92-06965A-13
23	45	100.0	798	4	US-08-867-611-36
24	45	100.0	1786	4	US-08-444-818-54
25	45	100.0	2261	4	US-08-444-818-66
26	45	100.0	2436	4	US-08-444-818-75
27	45	100.0	2772	4	US-08-444-818-89

28	45	100.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
29	45	100.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
30	45	100.0	2894	3	US-08-467-902A-23	Sequence 23, Appl
31	45	100.0	2894	4	US-09-275-265-23	Sequence 23, Appl
32	45	100.0	2955	2	US-08-443-260-3	Sequence 3, Appl
33	45	100.0	2955	3	US-08-442-805A-3	Sequence 3, Appl
34	45	100.0	2955	3	US-08-443-900A-3	Sequence 3, Appl
35	45	100.0	2955	4	US-08-444-818-124	Sequence 124, App
36	45	100.0	2955	4	US-08-249-843-3	Sequence 3, Appl
37	45	100.0	2995	4	US-08-444-818-138	Sequence 138, App
38	45	100.0	3011	1	US-08-188-281B-1	Sequence 1, Appl
39	45	100.0	3011	1	US-08-453-552-1	Sequence 1, Appl
40	45	100.0	3011	1	US-08-440-103-36	Sequence 36, Appl
41	45	100.0	3011	1	US-08-440-542-36	Sequence 36, Appl
42	45	100.0	3011	1	US-07-910-760-10	Sequence 10, Appl
43	45	100.0	3011	1	US-08-440-519-10	Sequence 10, Appl
44	45	100.0	3011	1	US-08-231-368-36	Sequence 36, Appl
45	45	100.0	3011	1	US-08-440-210-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-42
; Sequence 42, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silvert, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-214-650-42

Query Match 100.0%; Score 45; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ILDSFDPLV 9
|||||||

Db 1 ILDSFDPLV 9

RESULT 2

US-08-146-028-383

; Sequence 383, Application US/08146028

; Patent No. 5891640

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES

; CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR

; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED

; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,

; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM

; NUMBER OF SEQUENCES: 453

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/146,028

; INFORMATION FOR SEQ ID NO: 383:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-146-028-383

Query Match 100.0%; Score 45; DB 2; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9

Db 1 ILDSFDPLV 9

RESULT 3

US-08-723-425A-383

; Sequence 383, Application US/08723425A

; Patent No. 6165730

; GENERAL INFORMATION:

; APPLICANT: DELEYS, ROBERT

; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF

; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT

; TITLE OF INVENTION: EPITOPES AND THEIR USE IN A PROCESS FOR DETERMINATION OF

; TITLE OF INVENTION: ANTIBODIES OR BIOTINYLATED PEPTIDES CORRESPONDING ...

; NUMBER OF SEQUENCES: 453

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE, P.C.

; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

; CITY: Arlington

; STATE: VA

; COUNTRY: USA

; ZIP: 22201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/723,425A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: SADOFF, B.J.

; REGISTRATION NUMBER: 36,663

; REFERENCE/DOCKET NUMBER: 1487-13

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-816-4000

; TELEFAX: 703-816-4100

; INFORMATION FOR SEQ ID NO: 383:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-723-425A-383

Query Match 100.0%; Score 45; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9

Db 1 ILDSFDPLV 9

RESULT 4

US-09-112-206-383

; Sequence 383, Application US/09112206

; Patent No. 6210903

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES

; CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THE

; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED

; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES

; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM

; NUMBER OF SEQUENCES: 453

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/112,206

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/146,028

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 383:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-112-206-383

Query Match 100.0%; Score 45; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9

Db 1 ILDSFDPLV 9

RESULT 5

US-08-262-037-103

; Sequence 103, Application US/08262037

; Patent No. 5747239

; GENERAL INFORMATION:

; APPLICANT: Chang Yi Wang and Barbara Hosein

; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSTICS OF HCV

; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

; NUMBER OF SEQUENCES: 136


```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-103

```

```

Query Match 100.0%; Score 45; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
Db 21 ILDSFDPLV 29

RESULT 6
US-08-262-037-106
; Sequence 106, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-106

Query Match 100.0%; Score 45; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
Db 7 ILDSFDPLV 15

RESULT 7
US-08-262-037-5
; Sequence 5, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275

```

; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-5

Query Match 100.0%; Score 45; DB 1; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
| | | | | | | | | |
Db 30 ILDSFDPLV 38

RESULT 8
US-08-262-037-13
; Sequence 13, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; REGISTRATION NUMBER: 29,323

; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-13

Query Match 100.0%; Score 45; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
| | | | | | | | | |
Db 19 ILDSFDPLV 27

RESULT 9
US-08-262-037-104
; Sequence 104, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:

```
; TOPOLOGY: Unknown
; US-08-262-037-104

Query Match          100.0%; Score 45; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
   |||||
Db 36 ILDSFDPLV 44

RESULT 10
US-08-530-550-36
; Sequence 36, Application US/08530550
; Patent No. 5736321
; GENERAL INFORMATION:
; APPLICANT: Hosein, Barbara
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Peptides Effective for Diagnosis and
; TITLE OF INVENTION: Detection of Hepatitis c Infection
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 20002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)273-1717
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-530-550-36

Query Match          100.0%; Score 45; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
   |||||
Db 41 ILDSFDPLV 49

RESULT 11
US-08-262-037-12
; Sequence 12, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-12

Query Match          100.0%; Score 45; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
   |||||
Db 41 ILDSFDPLV 49

RESULT 12
US-08-444-818-40
; Sequence 40, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-40

Query Match 100.0%; Score 45; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
Db 53 ILDSFDPLV 61

RESULT 13
PCT-US94-04174-17
; Sequence 17, Application PC/TUS9404174
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: HCV Immunodiagnostic Antigens and Antibodies
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter J. Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-324-0880
; TELEFAX: 415-324-0960
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Clone 36-1 protein coding sequence
PCT-US94-04174-17

Query Match 100.0%; Score 45; DB 5; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
Db 134 ILDSFDPLV 142

RESULT 14
US-07-853-985A-6
; Sequence 6, Application US/07853985A
; Patent No. 5436318
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: P.O. BOX 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,985A
; FILING DATE: 19920320
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,611
; FILING DATE: 06-APR-1990
; APPLICATION NUMBER: US 07/594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Dehlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: 4600-0076.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-324-0880
; TELEFAX: 415-324-0960
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 504 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-853-985A-6

Query Match 100.0%; Score 45; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
Db 133 ILDSFDPLV 141

```

RESULT 15
US-07-681-703B-6
; Sequence 6, Application US/07681703B
; Patent No. 5443965
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/681,703B
; FILING DATE: 05-APR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-681-703B-6

```

```

Query Match      100.0%; Score 45; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ILDSFDPLV 9
   |
Db 133 ILDSFDPLV 141

```

Search completed: August 23, 2002, 09:57:53
Job time: 203 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:53 ; Search time 78.82 Seconds
(without alignments)
10.972 Million cell updates/sec

Title: US-08-854-825-42

Perfect score: 45

Sequence: 1 ILDSFDPLV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	100.0	663	2 E96705	hypothetical prote
2	45	100.0	3011	1 GNWVC3	genome polyprotein
3	45	100.0	3011	1 GNWVCH	genome polyprotein
4	45	100.0	3011	1 S40770	genome polyprotein
5	41	91.1	89	2 S10002	C8-2 protein - hep
6	41	91.1	103	2 S20763	nonstructural prot
7	41	91.1	173	2 PQ0251	polyprotein (clone
8	41	91.1	175	2 PQ0253	polyprotein (clone
9	41	91.1	182	2 PQ0252	polyprotein (clone
10	41	91.1	200	2 PQ0250	polyprotein (clone
11	41	91.1	3010	1 A45573	genome polyprotein
12	41	91.1	3010	1 GNWVTW	genome polyprotein
13	40	88.9	3010	1 GNWVTC	genome polyprotein
14	39	86.7	3010	1 GNWVCJ	genome polyprotein
15	37	82.2	153	2 PQ0255	polyprotein (clone
16	37	82.2	160	2 PQ0254	polyprotein (clone
17	37	82.2	200	2 PQ0248	polyprotein (clone
18	37	82.2	319	2 T46594	phytoene synthase
19	37	82.2	346	2 D84833	hypothetical prote
20	37	82.2	424	2 B84301	DNA primase [impor
21	37	82.2	2484	2 T26216	hypothetical prote
22	37	82.2	2607	2 T26215	hypothetical prote
23	37	82.2	3010	1 S18030	genome polyprotein
24	36	80.0	3033	1 J01303	genome polyprotein
25	34	75.6	208	2 E84316	hypothetical prote
26	34	75.6	286	2 D84591	hypothetical prote
27	34	75.6	300	2 AG3214	cultivar specific
28	34	75.6	354	2 F64325	aspartate-semialde
29	34	75.6	466	2 G72603	nitrate reductase

30	34	75.6	515	2 T08156	RNA maturase (EC 2
31	34	75.6	617	2 C95906	hypothetical prote
32	34	75.6	733	2 A97415	hypothetical prote
33	34	75.6	1876	2 T13801	phosphoinositide 3
34	33	73.3	96	2 E97868	hypothetical prote
35	33	73.3	137	2 E70653	hypothetical prote
36	33	73.3	265	2 I46986	albumin - dog (fra
37	33	73.3	278	2 T11550	probable membrane
38	33	73.3	318	2 C72856	AcOrf-51 protein -
39	33	73.3	319	2 T41795	ACMNPV orf51 - Bom
40	33	73.3	328	2 A84167	UDP-glucose 4-epim
41	33	73.3	357	2 F82878	XAA-PRO aminopepti
42	33	73.3	393	2 E90169	AAA family ATPase
43	33	73.3	409	2 F72504	probable 26S prote
44	33	73.3	452	2 T25076	hypothetical prote
45	33	73.3	521	2 T05944	cytochrome P450 93

ALIGNMENTS

RESULT 1

E96705

hypothetical protein T22E19.16 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E96705

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E96705

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-663 <STO>

A:Cross-references: GB:AE005173; NID:g6715731; PIDN:AAF26492.1; GSPDB:GN00141

C:Genetics:

A:Gene: T22E19.16

A:Map position: 1

Query Match 100.0%; Score 45; DB 2; Length 663;
Best Local Similarity 100.0%; Pred. NO. 0.96; DB 2; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Oy 1 ILDSFDPLV 9

|||||

Db 384 ILDSFDPLV 392

RESULT 2

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.H.
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: PQ0393; MUID:92268871
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DBJ:DI0128
 A:Experimental source: isolates E-b16
 A:Accession: PQ0404
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 2252 ILDSFDPLV 2260

RESULT 3
 GNWVCH
 genome polyprotein - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
 A:Reference number: A41546; MUID:92052256
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 2252 ILDSFDPLV 2260

RESULT 4
 S40770
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PCL285
 R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OKA>
 A:Cross-references: EMBL:DI0749; NID:g221586; PIDN:BA01582.1; PID:g221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PCL284; MUID:91013116
 A:Accession: PCL285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 45; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 2252 ILDSFDPLV 2260

RESULT 5
 S10002
 C8-2 protein - hepatitis C virus (fragment)
 C:Species: hepatitis C virus
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 17-Nov-2000
 C:Accession: S10002
 R:Maeno, M.; Kaminaka, K.; Sugimoto, H.; Esumi, M.; Hayashi, N.; Komatsu, K.; Abe, K.
 Nucleic Acids Res. 18, 2685-2689, 1990

A;Title: A cDNA clone closely associated with non-A, non-B hepatitis.

A;Reference number: S10002; MUID:90251443

A;Accession: S10002

A:Molecule type: DNA

A;Residues: 1-89 <MAX>

A;Cross-references: EMBL:X51587; NID:960687; PIDN:CAA35937.1; PID:g930250

C;Superfamily: hepatitis C virus genome polyprotein

Query Match 91.1%; Score 41; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8

|||||||

Db 76 ILDSFDPL 83

RESULT 6

S20763

nonstructural protein 5 - hepatitis C virus (fragment)

N;Alternate names: protein NS5

C;Species: hepatitis C virus

C;Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000

C;Accession: S20763

R;Murakawa, K.; Esumi, M.; Kato, K.; Kambara, H.; Shikata, T.

submitted to the EMBL Data Library, May 1991

A;Description: The heterogeneity of hepatitis C viruses isolated from a single patient an

A;Reference number: S20763

A;Accession: S20763

A:Molecule type: mRNA

A;Residues: 1-103 <MUR>

A;Cross-references: EMBL:X59609; NID:959488; PIDN:CAA42171.1; PID:g59489

A;Experimental source: isolate HCV33

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: nonstructural protein

Query Match 91.1%; Score 41; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8

|||||||

Db 91 ILDSFDPL 98

RESULT 7

PQ0251

polyprotein (clone 2258) - hepatitis C virus (isolate 22) (fragment)

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Nov-2000

C;Accession: PQ0251

R;Oshima, M.; Tsuchiya, M.; Yagasaki, M.; Orita, T.; Hasegawa, M.; Tomonoh, K.; Kojima,

J. Gen. Virol. 72, 2805-2809, 1991

A;Title: cDNA clones of Japanese hepatitis C virus genomes derived from a single patient

A;Reference number: PQ0245; MUID:92044457

A;Accession: PQ0251

A:Molecule type: genomic RNA

A;Residues: 1-173 <OSH>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 91.1%; Score 41; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8

|||||||

Db 133 ILDSFDPL 140

RESULT 8

PQ0253

polyprotein (clone 2220) - hepatitis C virus (isolate 22) (fragment)

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Nov-2000

C;Accession: PQ0253

R;Oshima, M.; Tsuchiya, M.; Yagasaki, M.; Orita, T.; Hasegawa, M.; Tomonoh, K.; Kojim

J. Gen. Virol. 72, 2805-2809, 1991

A;Title: cDNA clones of Japanese hepatitis C virus genomes derived from a single pati

A;Reference number: PQ0245; MUID:92044457

A;Accession: PQ0253

A:Molecule type: genomic RNA

A;Residues: 1-175 <OSH>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 91.1%; Score 41; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8

|||||||

Db 133 ILDSFDPL 140

RESULT 9

PQ0252

polyprotein (clone 2248) - hepatitis C virus (isolate 22) (fragment)

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Nov-2000

C;Accession: PQ0252

R;Oshima, M.; Tsuchiya, M.; Yagasaki, M.; Orita, T.; Hasegawa, M.; Tomonoh, K.; Kojim

J. Gen. Virol. 72, 2805-2809, 1991

A;Title: cDNA clones of Japanese hepatitis C virus genomes derived from a single pati

A;Reference number: PQ0245; MUID:92044457

A;Accession: PQ0252

A:Molecule type: genomic RNA

A;Residues: 1-182 <OSH>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 91.1%; Score 41; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8

|||||||

Db 133 ILDSFDPL 140

RESULT 10

PQ0250

polyprotein (clone 2206) - hepatitis C virus (isolate 22) (fragment)

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Nov-2000

C;Accession: PQ0250

R;Oshima, M.; Tsuchiya, M.; Yagasaki, M.; Orita, T.; Hasegawa, M.; Tomonoh, K.; Kojim

J. Gen. Virol. 72, 2805-2809, 1991

A;Title: cDNA clones of Japanese hepatitis C virus genomes derived from a single pati

A;Reference number: PQ0245; MUID:92044457

A;Accession: PQ0250

A:Molecule type: genomic RNA

A;Residues: 1-200 <OSH>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 91.1%; Score 41; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPL 8
Db 133 ILDSFDPL 140

RESULT 11
A45573
genome polypotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
A:Reference number: A45573; MUID:92295714
A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polypotein
C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 91.1%; Score 41; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPL 8
Db 2252 ILDSFDPL 2259

RESULT 12
GNWVTC
genome polypotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S. Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polypotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolyase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus (strain Taiwan)
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,
Query Match 91.1%; Score 41; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPL 8
Db 2252 ILDSFDPL 2259

RESULT 13
GNWVTC
genome polypotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E. J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from hum
A:Reference number: A38465; MUID:91140698
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAK>
A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polypotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolyase; nonstruct
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 88.9%; Score 40; DB 1; Length 3010;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPL 8
Db 2252 VLDSDPL 2259

RESULT 14
GNWVCU
genome polypotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstruct
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patien
A:Reference number: A39253; MUID:91088550

A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variability
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KA2>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine protease
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2253

Query Match 86.7%; Score 39; DB 1; Length 3010;
Best Local Similarity 87.5%; Pred. No. 80;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

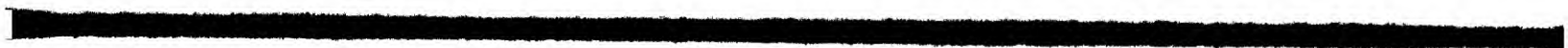
Qy 1 ILDSFDPL 8
| | | | | | | |
Db 2252 ILDSFDPI 2259

RESULT 15
P00255
Polyprotein (clone 2211) - hepatitis C virus (isolate 22) (fragment)
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Nov-2000
C:Accession: P00255
R:Oshima, M.; Tsuchiya, M.; Yagasaki, M.; Orita, T.; Hasegawa, M.; Tomonoh, K.; Kojima, J. Gen. Virol. 72, 2805-2809, 1991
A:Title: cDNA clones of Japanese hepatitis C virus genomes derived from a single patient
A:Reference number: P00245; MUID:92044457
A:Accession: P00255
A:Molecule type: genomic RNA
A:Residues: 1-153 <OSH>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 82.2%; Score 37; DB 2; Length 153;
Best Local Similarity 87.5%; Pred. No. 6.6;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8
| | | | | | | |
Db 133 ILDSFEPL 140

Search completed: August 23, 2002, 09:56:55
Job time: 145 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:27 ; Search time 37 Seconds
(without alignments)
9.418 Million cell updates/sec

Title: US-08-854-825-42

Perfect score: 45

Sequence: 1 ILDSFDPLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	45	100.0	3011	1	POLG_HCV1
2	45	100.0	3011	1	POLG_HCVH
3	41	91.1	3010	1	POLG_HCVHT
4	41	91.1	3010	1	POLG_HCVTW
5	40	88.9	3010	1	POLG_HCVBK
6	39	86.7	3010	1	POLG_HCVJA
7	37	82.2	424	1	PRIM_HALN1
8	36	80.0	3033	1	POLG_HCVJ6
9	34	75.6	354	1	DHAS_METJA
10	34	75.6	515	1	MATK_PICGL
11	34	75.6	515	1	MATK_PICMA
12	34	75.6	515	1	MATK_PICRU
13	34	75.6	608	1	ALBU_RABIT
14	33	73.3	278	1	P29K_STRGC
15	33	73.3	290	1	FAC2_DROME
16	33	73.3	318	1	Y051_NPVAC
17	33	73.3	519	1	IRX4_HUMAN
18	33	73.3	588	1	PEOB_SOLTU
19	33	73.3	608	1	ALBU_CANFA
20	33	73.3	845	1	RRPO_IPNVJ
21	32	71.1	139	1	UGRI_MOUSE
22	32	71.1	387	1	MSOX_BACSN
23	32	71.1	414	1	YC44_GUITH
24	32	71.1	485	1	GATA_BACSU
25	32	71.1	493	1	SYK_ARCFU
26	32	71.1	675	1	UVRB_NEIGO
27	32	71.1	675	1	UVRB_NEIMA
28	32	71.1	675	1	UVRB_NEIMB
29	32	71.1	712	1	SLM1_TRICF
30	31	68.9	103	1	LSM7_HUMAN
31	31	68.9	160	1	VD14_BPT5
32	31	68.9	237	1	LECA_DOLLA
33	31	68.9	267	1	DMAL_SCHPO

RESULT 1

ID	POLG_HCV1	STANDARD	PRT	3011 AA
AC	P26664			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the p6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -;			
DR	PIR; A39166; GNMVC3.			
DR	HSSP; P27958; LHEI.			
DR	MEROPS; S29.001; -;			
DR	MEROPS; U39.001; -;			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002531; HCV_NS1.			
DR	InterPro; IPR002518; HCV_NS2.			
DR	InterPro; IPR004109; HCV_NS3.			
DR	InterPro; IPR000745; HCV_NS4a.			

34	31	68.9	353	1	HMD_METTL
35	31	68.9	453	1	YAW6_SCHPO
36	31	68.9	467	1	RFAL_CRIFA
37	31	68.9	564	1	SVT_MYCGE
38	31	68.9	601	1	YE06_SCHPO
39	31	68.9	668	1	MTMW_METWO
40	31	68.9	843	1	RA25_YEAST
41	31	68.9	986	1	DP01_SYNY3
42	31	68.9	1063	1	YNH6_CABEL
43	31	68.9	4273	1	PKSM_BACSU
44	30	66.7	150	1	CB22_PINSY
45	30	66.7	174	1	CET1_TOBAC

ALIGNMENTS

DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00492; HELIC3; 1.
KW Core protein; Glycoprotein; Transfrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT CHAIN 3012 369
FT TRANSMEM 347 369
FT ACT_SITE 1083 1093
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
|||||||
Db 2252 ILDSFDPLV 2360

RESULT 2

POIG_HCVH STANDARD; PRT; 3011 AA.
ID POLG_HCVH
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M67463; AAA45534.1; -.
CC PIR; A36814; GMYVCH.
CC PDB; 1HEI; 25-NOV-98.
CC PDB; 1A1V; 16-FEB-99.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.

DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF01560; HCV_NS2; 1.
DR Pfam: PF01538; HCV_NS3; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PD186062; HCV_NS1; 1.
DR SMART; SM00492; HelicC3; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
| | | | | | | | | |
Db 2252 ILDSFDPLV 2260

RESULT 3
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC 000269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_taxid=31642;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11168; BAA01943.1; -
CC PIR; A45573; A45573.
CC HSP; P26663; IJXP.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro: IPR001410; DEAD
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00492; HELIC3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1
 FT CHAIN 115
 FT CHAIN 191
 FT CHAIN 384
 FT CHAIN 729
 FT CHAIN 1006
 FT CHAIN 1615
 FT CHAIN 1662
 FT CHAIN 1863
 FT CHAIN 2013
 FT CHAIN 3010
 FT TRANSMEM 347
 FT ACT_SITE 1083
 FT ACT_SITE 1107
 FT ACT_SITE 1165
 FT NP_BIND 1230
 FT SITE 1316
 FT CARBOHYD 196
 FT CARBOHYD 209
 FT CARBOHYD 234
 FT CARBOHYD 250
 FT CARBOHYD 305
 FT CARBOHYD 417
 FT CARBOHYD 423
 FT CARBOHYD 430
 FT CARBOHYD 448
 FT CARBOHYD 532
 FT CARBOHYD 540
 FT CARBOHYD 556
 FT CARBOHYD 576
 FT CARBOHYD 623
 FT CARBOHYD 645
 FT CARBOHYD 2041
 FT CARBOHYD 2077
 FT CARBOHYD 2240
 FT CARBOHYD 2329
 FT CARBOHYD 2788
 FT CARBOHYD 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 91.1%; Score 41; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ILDSFDPL 8
 |||||
 Db 2252 ILDSFDPL 2259

RESULT 4
 ID POLG_HCVTW
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RL mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC !- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC !- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC !- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC !- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84754; -; NOT_ANNOTATED_CDS.
 CC PIR; A40244; GNVVTW.
 CC HSSP; P26663; LUXP.
 CC MEROPS; S29.001; -;
 CC MEROPS; U39.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1
 FT CHAIN 115
 FT CHAIN 191
 FT CHAIN 384
 FT CHAIN 729
 FT CHAIN 1006
 FT CHAIN 1615
 FT CHAIN 1662
 FT CHAIN 1863
 FT CHAIN 2013
 FT CHAIN 3010
 FT TRANSMEM 347

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDDE215 CRC64;

 Query Match 91.1%; Score 41; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 ILDSFDPL 8
 DB 2252 ILDSFDPL 2259

 RESULT 5
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RN SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (2.4 'ANGSTROMS) OF 1027-1215.

RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (2.2 'ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M58335; AAA72945.1; -;
 DR PIR; A38465; GNMVTC.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR002522; HCV_core.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RdRP; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

```

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 88.9%; Score 40; DB 1; Length 3010;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILDSFDPL 8
DB 2252 VLDSFDPL 2259
:|||||||

RESULT 6
POLG.HCVJA STANDARD; PRT; 3010 AA.
AC P26562;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.

```

```

RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijioka M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D90208; BAA14233.1; .
CC PIR; A39253; GNWVCJ.
CC HSSP; P26663; 1JXP.
CC MEROPS; S29.001; .
CC MEROPS; U39.001; .
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRP.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00998; HCV_RdRP; 1.
CC Pfam; PF00271; helicase_C; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00492; HelicC3; 1.
KW Polyprotein; Glycoprotein;
KW Core protein; Coat protein;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165

```

Transferase; RNA-directed RNA polymerase;
Envelope protein; Helicase; ATP-binding;
protein; Hydrolase; Serine protease.
REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
CAPSID PROTEIN C (POTENTIAL).
MATRIX PROTEIN (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).
NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
PROTEASE/HELICASE NS3 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
POTENTIAL.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).

```
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 86.7%; Score 39; DB 1; Length 3010;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPL 8
Db 2252 ILDSFDPI 2259

RESULT 7
PRIM_HA1N1
ID PRIM_HA1N1 STANDARD; PRT; 424 AA.
AC QHP03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative DNA primase (EC 2.7.7.-).
GN PRI OR VNG1470G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Kellier K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.F., Krebs M.P., Angevine C.M., Dale H.,
RA Iisenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RC Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL RNA
CC PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC PRIMASE SMALL SUBUNIT
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
```

```
DR EMBL; AE005062; AAG19774.1; -.
DR InterPro; IPR002755; DNA_primase_S.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF01896; DNA_primase_S; 1.
DR SMART; SM00278; Hhh1; 1.
DR KW Hypothetical protein; Transferase; DNA replication;
DR KW DNA-directed RNA polymerase; Primosome; Metal-binding;
DR KW Complete proteome.
FT ACT_SITE 131 131 POTENTIAL.
FT ACT_SITE 133 133 POTENTIAL.
SQ SEQUENCE 424 AA; 47174 MW; 946BDEF096E1DCAD CRC64;

Query Match 82.2%; Score 37; DB 1; Length 424;
Best Local Similarity 87.5%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LDSFDPLV 9
Db 360 LDSFDPLV 367

RESULT 8
POLG_HCVJ6
ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RC J. Gen. Virol. 72:2697-2704(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; D00944; BAA00792.1; -.
DR PIR; J01303; J01303.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
-----
```

```
DR MEROPS: U39_001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RDRP; 1.
DR Pfam: PF00271; helicase_C; 1.
DR ProDom: PD166062; HCV_NS1; 1.
DR PolyProtein: Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
```

Query Match 80.0%; Score 36; DB 1; Length 3033;
Best Local Similarity 66.7%; Pred. No. 1; lfe-02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
:11111111
DB 2252 VLDSLDPMV 2260

```
RESULT 9
DHAS_METJA STANDARD; PRT; 354 AA.
AC Q57658;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA
DE dehydrogenase) (ASADH).
GN ASD_OR MJ0205.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A. 2561 / ATCC 43067;
RC STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Bult C.J., White O., Olsen G.J., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kervilange A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.D., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE SECOND STEP IN THE COMMON
CC METABOLIC PATHWAY TO SYNTHESIZE THR AND MET FROM ASPARTIC ACID.
CC -!- CATALYTIC ACTIVITY: L-aspartate-4-semialdehyde + phosphate +
CC NADP(+) = L-4-aspartyl phosphate + NADPH.
CC -!- PATHWAY: METHIONINE BIOSYNTHESIS, THREONINE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE ASPARTATE-SEMIALDEHYDE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67476; AAB98189.1; -
CC TIGR: MJ0205; -
CC InterPro: IPR000319; Asp_semiald_dh.
CC InterPro: IPR000534; Semialdh_dh.
CC Pfam: PF01118; Semialdh_dh; 1.
CC Pfam: PF02774; Semialdh_dh; 1.
CC PROSITE: PS01103; ASD; 1.
CC Oxidoreductase; NADP; Methionine biosynthesis; Threonine biosynthesis;
KW Amino-acid biosynthesis; Complete proteome.
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 354 AA; 394339 MW; 5C412CA5C39B166E CRC64;
```

Query Match 75.6%; Score 34; DB 1; Length 354;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILDSFDPL 8
:11111111
DB 270 VMDKFDPL 277

RESULT 10
MATK_PICGL STANDARD; PRT; 515 AA.
ID MATK_PICGL
AC O63070;

```

OC Spermatophyta; Coniferales; Pinaceae; Pinaceae; Picea.
NCBI_TaxID=3335;
[1]
RN SEQUENCE FROM N.A.
RA Germano J., Klein A.S.;
RT "Distribution of sequence variation in the chloroplast trnK intron and
RT rpl33-psaJ-trnp intergenic spacers within and between Picea rubens,
RT Picea mariana and Picea glauca: potential molecular markers for
RT species-specific identification.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF059343; AAC14710.1; -.
CC DR InterPro; IPR000442; Intron_maturse2.
CC DR InterPro; IPR002866; MatK_N.
CC DR Pfam; PF01348; Intron_maturas2; 1.
CC DR Pfam; PF01824; MatK_N; 1.
CC DR Kew Chloroplast; mRNA processing.
CC KW CHLOROPLAST; mRNA processing.
CC SQ SEQUENCE 515 AA; 60951 MW; 6DAA84436C433F99 CRC64;

Query Match 75.6% Score 34; DB 1; Length 515;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
   | | | | |
Db 356 ITDEFDPIV 364

RESULT 12
MATK_PICRU
ID MATK_PICRU STANDARD; PRT; 515 AA.
AC O63071;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK OR YCF14.
OS Picea rubens (Red spruce).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferales; Pinaceae; Picea.
OC NCBI_TaxID=3333;
[1]
RN SEQUENCE FROM N.A.
RA Germano J., Klein A.S.;
RT "Distribution of sequence variation in the chloroplast trnK intron and
RT rpl33-psaJ-trnp intergenic spacers within and between Picea rubens,
RT Picea mariana and Picea glauca: potential molecular markers for
RT species-specific identification.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AF059342; AAC14709.1; -
 DR InterPro: IPR000442; Intron_mature2.
 DR InterPro: IPR002866; MatK_N.
 DR Pfam: PF01348; Intron_matures2; 1.
 DR Pfam: PF01824; MatK_N; 1.
 KW Chloroplast; mRNA processing.
 SQ SEQUENCE 515 AA; 60951 MW; 001E940678127A89 CRC64;

Query Match 75.6%; Score 34; DB 1; Length 515;
 Best Local Similarity 66.7%; Pred. No. 38;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 I I I I I I
 DB 356 ITDFDPIV 364

RESULT 13
 ALBU_RABIT STANDARD; PRT; 608 AA.
 AC P49065;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Serum albumin precursor.
 GN ALB.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
 RA Sheffield W.P., Syed S., Schuyler P.D.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: SERUM ALBUMIN. THE MAIN PROTEIN OF PLASMA, HAS A GOOD
 CC BINDING CAPACITY FOR WATER, CA++, NA+, K+, FATTY ACIDS, HORMONES,
 CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
 CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: PLASMA.
 CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U18344; AAB58347.1; -
 DR HSP; P02768; 1BJ5.
 DR InterPro: IPR000264; Serum_albumin.
 DR Pfam: PF00273; transport_prot; 3.
 DR PRINTS: PR00802; SERUMALBUMIN.
 DR ProDom: PD002486; Serum_albumin; 1.
 DR SMART: SM00103; ALBUMIN; 3.
 DR PROSITE: PS00212; ALBUMIN; 3.
 KW Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;
 KW Copper.

FT SIGNAL 1 18 BY SIMILARITY.
 FT PROPEP 19 24 BY SIMILARITY.
 FT CHAIN 25 608 SERUM ALBUMIN.
 FT REPEAT 28 202 1.
 FT REPEAT 221 394 2.

FT REPEAT 413 592 3.
 FT METAL 27 27 COPPER.
 FT BY SIMILARITY.
 FT DISULFID 77 86 BY SIMILARITY.
 FT DISULFID 99 115 BY SIMILARITY.
 FT DISULFID 114 125 BY SIMILARITY.
 FT DISULFID 148 193 BY SIMILARITY.
 FT DISULFID 192 201 BY SIMILARITY.
 FT DISULFID 224 270 BY SIMILARITY.
 FT DISULFID 269 277 BY SIMILARITY.
 FT DISULFID 289 303 BY SIMILARITY.
 FT DISULFID 302 313 BY SIMILARITY.
 FT DISULFID 340 385 BY SIMILARITY.
 FT DISULFID 384 393 BY SIMILARITY.
 FT DISULFID 416 462 BY SIMILARITY.
 FT DISULFID 461 472 BY SIMILARITY.
 FT DISULFID 485 501 BY SIMILARITY.
 FT DISULFID 500 511 BY SIMILARITY.
 FT DISULFID 538 583 BY SIMILARITY.
 FT DISULFID 582 591 BY SIMILARITY.
 SQ SEQUENCE 608 AA; 68914 MW; CF5E92647AAFE9A2 CRC64;

Query Match 75.6%; Score 34; DB 1; Length 608;
 Best Local Similarity 66.7%; Pred. No. 46;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 I I I I I I
 DB 397 VLDEFQPLV 405

RESULT 14
 P29K_STRGC STANDARD; PRT; 278 AA.
 ID P29K_STRGC
 AC P42361;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 29 kDa membrane protein in psaa 5'region (ORF1).
 OS Streptococcus gordonii Challis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID=29390;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=PK488;
 RX MEDLINE=95012638; PubMed=7927711;
 RA Kolenbrander P.E., Andersen R.N., Ganeshkumar N.;
 RT "Nucleotide sequence of the Streptococcus gordonii PK488
 coaggregation adhesin gene, scaA, and ATP-binding cassette.";
 RL Infect. Immun. 62:4469-4480(1994).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE
 CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: L11577; AAA71946.1; -
 DR InterPro: IPR001626; ABC-3.
 DR Pfam: PF00950; ABC-3; 1.
 KW Transmembrane; Transport.
 FT TRANSMEM 18 38 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 134 154 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.

FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 246 266 POTENTIAL.
 SQ SEQUENCE 278 AA; 29705 MW; 47C1117C97DE084C CRC64;

Query Match 73.3%; Score 33; DB 1; Length 278;
 Best Local Similarity 66.7%; Pred. No. 30;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
 Db 158 LITSFDPLV 166

RESULT 15
 FAC2_DROME STANDARD; PRT; 290 AA.
 AC Q9U1H8; Q9VRM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE CAX prenyl protease 2 (EC 3.4.22.-) (Frenyl protein-specific
 DE endoprotease 2) (Farnesylated-proteins converting enzyme 2) (FACE-2)
 DE (Severas protein).
 GN SRAS OR CG4852.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Weinzierl-Hinun A., Toeroek I., Kiss I., Farkas R., Mechler B.M.;
 RT "the severas gene of Drosophila encodes a CAX-protease and acts as a
 RT tumour suppressor."; submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fogle C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED AND GERANYLATED PROTEINS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ252068; CAB64383.1;
 DR EMBL; AE003565; AAF50770.2;
 DR Flybase; FBgn0029121; Sras.
 DR InterPro; IPR003675; ABl.
 DR Pfam; PF02517; ABl; 1.
 DR HydroLase; Thiol protease; Transmembrane; Endoplasmic reticulum.
 KW TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 159 179 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT CONFLICT 82 83 KL -> MV (IN REF. 1).
 FT CONFLICT 110 110 I -> M (IN REF. 1).
 FT CONFLICT 126 126 D -> H (IN REF. 1).
 SQ SEQUENCE 290 AA; 33082 MW; 157CAAA5B30693A8 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 290;
 Best Local Similarity 77.8%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9
 Db 160 ILQSFSPVLV 168

Search completed: August 23, 2002, 09:55:29
 Job time: 59 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:22:01 ; Search time 139.83 Seconds
(without alignments)
11.135 Million cell updates/sec

Title: US-08-854-825-42
Perfect score: 45
Sequence: 1 ILDSFDPLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	100.0	66	12	Q91DB5	Q91db5 hepatitis c
2	45	100.0	66	12	Q91DB4	Q91db4 hepatitis c
3	45	100.0	66	12	Q91DB2	Q91db2 hepatitis c
4	45	100.0	66	12	Q91DB1	Q91db1 hepatitis c
5	45	100.0	66	12	Q91DB0	Q91db0 hepatitis c
6	45	100.0	66	12	Q91DA9	Q91da9 hepatitis c
7	45	100.0	66	12	Q91DA8	Q91da8 hepatitis c
8	45	100.0	66	12	Q91DA7	Q91da7 hepatitis c
9	45	100.0	66	12	Q91DA6	Q91da6 hepatitis c
10	45	100.0	66	12	Q91DA5	Q91da5 hepatitis c
11	45	100.0	66	12	Q91DA4	Q91da4 hepatitis c
12	45	100.0	66	12	Q91DA3	Q91da3 hepatitis c
13	45	100.0	148	12	Q91SD7	Q91sd7 hepatitis c
14	45	100.0	448	12	Q9E3E3	Q9e3e3 hepatitis c
15	45	100.0	448	12	Q9E3E2	Q9e3e2 hepatitis c
16	45	100.0	448	12	Q9E3E1	Q9e3e1 hepatitis c

ALIGNMENTS

RESULT 1

Q91DB5 ID Q91DB5 PRELIMINARY; PRT; 66 AA.
AC Q91DB5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO.201;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R., Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity- Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864(2001).
DR EMBL; AB056538; BAB63170.1; --
FT NON_TER 1
FT NON_TER 66
SQ SEQUENCE 66 AA; 7360 MW; 8D9CCD3453391865 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILDSFDPLV 9

Db 44 ILDSFDPLV 52

RESULT 2

Q91DB4

ID Q91DB4 PRELIMINARY; PRT; 66 AA.
 AC Q91DB4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMPLE NO.11N;
 RX MEDLINE=21538940; PubMed=11682498;
 RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
 RA Fujita T., Oka K., Hotta H.;
 RT "Correlation between Mutations in the Interferon Sensitivity-
 RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
 RT Subtypes 1b, 1c, and 2a.";
 RL J. Clin. Microbiol. 39:3858-3864(2001).
 DR EMBL; AB056539; BAB63171.1; -.
 FT NON_TER 1
 FT NON_TER 66
 SQ SEQUENCE 66 AA; 7422 MW; 84A6DE1D6939187B CRC64;

Query Match 100.0%; Score 45; DB 12; Length 66;
 Best Local Similarity 100.0%; Pred. No. 0.088;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 44 ILDSFDPLV 52

RESULT 3
 Q91DB2 PRELIMINARY; PRT; 66 AA.
 ID Q91DB2;
 AC Q91DB2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMPLE NO.50;
 RX MEDLINE=21538940; PubMed=11682498;
 RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
 RA Fujita T., Oka K., Hotta H.;
 RT "Correlation between Mutations in the Interferon Sensitivity-
 RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
 RT Subtypes 1b, 1c, and 2a.";
 RL J. Clin. Microbiol. 39:3858-3864(2001).
 DR EMBL; AB056541; BAB63173.1; -.
 FT NON_TER 1
 FT NON_TER 66
 SQ SEQUENCE 66 AA; 7346 MW; 2D9DCC34533919A3 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 66;
 Best Local Similarity 100.0%; Pred. No. 0.088;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 44 ILDSFDPLV 52

RESULT 4

Q91DB1 PRELIMINARY; PRT; 66 AA.
 ID Q91DB1;
 AC Q91DB1;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMPLE NO.330;
 RX MEDLINE=21538940; PubMed=11682498;
 RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
 RA Fujita T., Oka K., Hotta H.;
 RT "Correlation between Mutations in the Interferon Sensitivity-
 RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
 RT Subtypes 1b, 1c, and 2a.";
 RL J. Clin. Microbiol. 39:3858-3864(2001).
 DR EMBL; AB056542; BAB63174.1; -.
 FT NON_TER 1
 FT NON_TER 66
 SQ SEQUENCE 66 AA; 7346 MW; 2D9DCC34533919A3 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 66;
 Best Local Similarity 100.0%; Pred. No. 0.088;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 44 ILDSFDPLV 52

RESULT 5
 Q91DB0 PRELIMINARY; PRT; 66 AA.
 ID Q91DB0;
 AC Q91DB0;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMPLE NO.130;
 RX MEDLINE=21538940; PubMed=11682498;
 RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
 RA Fujita T., Oka K., Hotta H.;
 RT "Correlation between Mutations in the Interferon Sensitivity-
 RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
 RT Subtypes 1b, 1c, and 2a.";
 RL J. Clin. Microbiol. 39:3858-3864(2001).
 DR EMBL; AB056543; BAB63175.1; -.
 FT NON_TER 1
 FT NON_TER 66
 SQ SEQUENCE 66 AA; 7346 MW; 2D9DCC34533919A3 CRC64;

Query Match 100.0%; Score 45; DB 12; Length 66;
 Best Local Similarity 100.0%; Pred. No. 0.088;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILDSFDPLV 9
 Db 44 ILDSFDPLV 52

```
RESULT 6
Q91DA9          PRELIMINARY;      PRT;      66 AA.
ID Q91DA9;
AC Q91DA9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO.85;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
RA Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity-
RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
RT Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864(2001).
DR EMBL; AB056544; BAB63176.1; -.
FT NON_TER 1
FT NON_TER 66
SQ SEQUENCE 66 AA; 7374 MW; 2D8AEB04533900BA CRC64;
```

```
Query Match          100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 ILDSFDPLV 9
   | | | | | | | |
Db 44 ILDSFDPLV 52
```

```
RESULT 7
Q91DA8          PRELIMINARY;      PRT;      66 AA.
ID Q91DA8;
AC Q91DA8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO.128;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
RA Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity-
RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
RT Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864(2001).
DR EMBL; AB056545; BAB63177.1; -.
FT NON_TER 1
FT NON_TER 66
SQ SEQUENCE 66 AA; 7346 MW; 2D9DCC34533919A3 CRC64;
```

```
Query Match          100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 ILDSFDPLV 9
   | | | | | | | |
Db 44 ILDSFDPLV 52
```

```
RESULT 8
Q91DA7          PRELIMINARY;      PRT;      66 AA.
ID Q91DA7;
AC Q91DA7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO.220;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
RA Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity-
RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
RT Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864(2001).
DR EMBL; AB056546; BAB63178.1; -.
FT NON_TER 1
FT NON_TER 66
SQ SEQUENCE 66 AA; 7346 MW; 2D9DCC34533919A3 CRC64;
```

```
Query Match          100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 ILDSFDPLV 9
   | | | | | | | |
Db 44 ILDSFDPLV 52
```

```
RESULT 9
Q91DA6          PRELIMINARY;      PRT;      66 AA.
ID Q91DA6;
AC Q91DA6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO.193;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,
RA Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity-
RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
RT Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864(2001).
DR EMBL; AB056547; BAB63179.1; -.
FT NON_TER 1
FT NON_TER 66
SQ SEQUENCE 66 AA; 7372 MW; 3C8CCC25423919A3 CRC64;
```

```
Query Match          100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 ILDSFDPLV 9
   | | | | | | | |
Db 44 ILDSFDPLV 52
```

D**b** 44 ILDSFDPLV 52

RESULT#	12
ID	Q91DA3
AC	Q91DA3:
DT	01-DEC-2001 (TTEM)
DT	01-DEC-2001 (TTEM)
DE	01-DEC-2001 (TTEM)
OS	POLYPROTEIN (FRAG)
OC	Hepatitis C virus
OC	Hepacivirus; SSRNA pos
RN	Vesiculovirus.
RX	NCBI_TaxID=11103;
RP	[1]
RC	SEQUENCE FROM N.A
RC	STRAIN=SAMPLE NO.
RL	MEDLINE=21538940;
RA	Lusida M.I., Nagai
RA	Fujita T., Oka K.
RT	"Correlation betw
RT	Determining Regio
RL	Subtypes lb, lc,
RL	J. Clin. Microbio
DR	EMBL; AB056550; B
FT	NON_TER
FT	NON_TER
SQ	SEQUENCE 66 AA;

```

Query Match
Best Local Similarity
Matches 9; Conserved
QY 1 ILDSFDPLV 9
      |||||
Db 44 ILDSFDPLV 52

```

RESULT	Q15D7	ID	Q91SD7	PRELIM
AC	Q91SD7	AC	Q91SD7	PRELIM
DT	01-DEC-2001	DT	01-DEC-2001	(TEMP)
DT	01-DEC-2001	DT	01-DEC-2001	(TEMP)
DE	01-DEC-2001	DE	01-DEC-2001	(TEMP)
DE	01-DEC-2001	DE	01-DEC-2001	(TEMP)
OS	NS5A PROTEIN	OS	NS5A PROTEIN	(FRACTION)
OS	Hepatitis C virus	OS	Hepatitis C virus	
OS	viruses; ssrna pos	OS	viruses; ssrna pos	
OC	Hepacivirus.	OC	Hepacivirus.	
ON	NCBI_TaxID=11103;	ON	NCBI_TaxID=11103;	
RP	[1]	RP	[1]	
RC	SEQUENCE FROM N.A.	RC	SEQUENCE FROM N.A.	
RA	STRAIN=42 MONTHS;	RA	STRAIN=42 MONTHS;	
RT	Rott V.;	RT	Rott V.;	
RT	"Evolution of diffe	RT	"Evolution of diffe	
RL	of a chronically i	RL	of a chronically i	
RL	Submitted (MAR-20	RL	Submitted (MAR-20	
FT	EMBL; AF359354; A	FT	EMBL; AF359354; A	
FR	NON_TER	FR	NON_TER	1
FR	NON_TER	FR	NON_TER	148
SQ	SEQUENCE	SQ	SEQUENCE	148 AA

		Query Match
		Best Local Similarity
		Matches 9; Conserved
Qy	1	ILDSFDPLV 9
Dd	103	ILDSFDPLV 111

```

RESULT 10
Q911DA5
ID Q911DA5 PRELIMINARY; PRT; 66 AA.
AC Q91DA5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMPLE NO. 324;
RX MEDLINE=21538940; PubMed=11682498;
RA Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipcto, Handajani R.,
RA Fujita T., Oka K., Hotta H.;
RT "Correlation between Mutations in the Interferon Sensitivity-
RT Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus
RT Subtypes 1b, 1c, and 2a.";
RL J. Clin. Microbiol. 39:3858-3864 (2001).
DR ENBL; AB056548; BAB63180.1; -.
FT NON_TER 1
FT TER 1
FT SEQUENCE 66 AA; 7365 MW; 3787CC2E493919A3 CRC64;
SQ

```

```

Query Match      100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ILDSFDPLV 9
Db       44  ILDSFDPLV 52

```

RESULT		11	
Q91DA4			
ID	Q91DA4	PRELIMINARY;	PRT; 66 AA.
AC	Q91DA4;		
DT	01-DEC-2001	(TrEMBLrel. 19,	Created)
DT	01-DEC-2001	(TrEMBRel. 19,	Last sequence update)
DT	01-DEC-2001	(TrEMBRel. 19,	Last annotation update)
DE	POLYPROTEIN	(FRAGMENT).	
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepadnavirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SAMPLE NO.295;		
RX	MEDLINE=21538940; PubMed=11682498;		
RA	Lusida M.I., Nagano-Fujii M., Nidom C.A., Soetjipto, Handajani R.,		
RA	Fujita T., Oka K., Hotta H.;		
RT	"Correlation between Mutations in the Interferon Sensitivity-		
RT	Determining Region of NS5A Protein and Viral Load of Hepatitis C Virus		
RT	Subtypes lb, lc, and 2a.";		
RL	J. Clin. Microbiol. 39:3858-3864(2001)..		
DR	EMBL: AB056549; BAB63181.1; -;		
FT	NON_TER	1	
FF	NON_TERM	i	
		66	
SQ	SEQUENCE	66 AA; 7296 MW;	109FDH8D505779B9AF CRC64;

```

Query Match      100.0%; Score 45; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. NO. 0.088;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ILDSFDPLV 9
          | | | | | | | |

```

```
RESULT 14
Q9E3E3 PRELIMINARY; PRT; 448 AA.
AC Q9E3E3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 3;
RX MEDLINE=20438098; PubMed=10982347;
RA Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA Carithers R.L. Jr., Gretch D.R.;
RT "Prospective characterization of full-length hepatitis C virus NS5A
RT quasisppecies during induction and combination antiviral therapy.";
RL J. Virol. 74:9028-9038(2000).
DR EMBL: AF264995; AAG21135.1; -.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase.
FT NON_TER 1
FT NON_TER 448
SQ SEQUENCE 448 AA; 48943 MW; 2E8F9EDB92E6BFAP CRC64;
```

Query Match 100.0%; Score 45; DB 12; Length 448;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ILDSFDPLV 9
    |
Db 280 ILDSFDPLV 288
```

```
RESULT 15
Q9E3E2 PRELIMINARY; PRT; 448 AA.
AC Q9E3E2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOME POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus type 1a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31646;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 3;
RX MEDLINE=20438098; PubMed=10982347;
RA Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA Carithers R.L. Jr., Gretch D.R.;
RT "Prospective characterization of full-length hepatitis C virus NS5A
RT quasisppecies during induction and combination antiviral therapy.";
RL J. Virol. 74:9028-9038(2000).
DR EMBL: AF264996; AAG21136.1; -.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase.
FT NON_TER 1
FT NON_TER 448
SQ SEQUENCE 448 AA; 49027 MW; 55083E02666F8DBE CRC64;
```

Query Match 100.0%; Score 45; DB 12; Length 448;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ILDSFDPLV 9
    |
Db 280 ILDSFDPLV 288
```

Search completed: August 23, 2002, 10:22:01
Job time: 1501 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 10:01:00 ; Search time 179.72 seconds
(without alignments)
6.180 Million cell updates/sec

Title: US-08-854-825-35
Perfect score: 55
Sequence: 1 LFNILGGWV 10
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	10	15	AA1980
2	55	100.0	10	15	AA1981
3	55	100.0	10	16	AA1982
4	55	100.0	10	16	AA1983
5	55	100.0	10	20	AA1984
6	55	100.0	10	20	AA1985
7	55	100.0	10	21	AA1986
8	55	100.0	10	22	AA1987
9	55	100.0	10	22	AA1988
10	55	100.0	10	22	AA1989
11	55	100.0	10	22	AA1990

12	55	100.0	10	22	AA1991
13	55	100.0	10	22	AA1992
14	55	100.0	10	22	AA1993
15	55	100.0	11	22	AA1994
16	55	100.0	11	22	AA1995
17	55	100.0	11	22	AA1996
18	55	100.0	11	22	AA1997
19	55	100.0	11	22	AA1998
20	55	100.0	15	22	AA1999
21	55	100.0	15	22	AA2000
22	55	100.0	20	22	AA2001
23	55	100.0	20	22	AA2002
24	55	100.0	138	10	AA2003
25	55	100.0	138	10	AA2004
26	55	100.0	140	13	AA2005
27	55	100.0	140	13	AA2006
28	55	100.0	166	13	AA2007
29	55	100.0	269	20	AA2008
30	55	100.0	293	17	AA2009
31	55	100.0	320	13	AA2010
32	55	100.0	320	13	AA2011
33	55	100.0	363	13	AA2012
34	55	100.0	363	17	AA2013
35	55	100.0	382	10	AA2014
36	55	100.0	382	10	AA2015
37	55	100.0	460	10	AA2016
38	55	100.0	460	10	AA2017
39	55	100.0	477	13	AA2018
40	55	100.0	477	13	AA2019
41	55	100.0	477	13	AA2020
42	55	100.0	477	13	AA2021
43	55	100.0	512	22	AA2022
44	55	100.0	592	14	AA2023
45	55	100.0	592	22	AA2024

ALIGNMENTS

RESULT 1
AA1980
ID AA1980 standard; peptide; 10 AA.
XX
AC AA1980
DT 11-MAY-1995 (first entry)
XX
DE Peptide fragment (1.0891) of HCV binds HLA-A2.1.
XX
KW antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
KW human immunodeficiency virus; human papilloma virus; p53; c-ERB2;
KW MAGE-1; melanoma antigen-1; core antigen; surface antigen;
KW pharmaceutical composition; in vivo; ex vivo; therapeutic;
KW diagnostic; MHC class I molecule; major histocompatibility complex;
KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen.
XX
OS Hepatitis C virus.
XX
FN WO9420127-A.
XX
PD 15-SEP-1994.
XX
PF 04-MAR-1994; 94WO-US02353.
XX
PR 05-MAR-1993; 93US-0027146.
PR 04-JUN-1993; 93US-0073205.
PR 29-NOV-1993; 93US-0159184.
XX
PA (CYTE-) CYTEL CORP.
XX
PI Grey HM, Kast WM, Sette A, Sidney J;
XX

DR WPI; 1994-302678/37.
 PT Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
 PT for treatment or prophylaxis of cancer, virus infection or
 PT autoimmune diseases.
 XX
 XX
 PS Example 5; Page 107; 138pp; English.
 XX
 CC AAR59496-R61666 are immunogenic 10mer peptides that contain a HLA-A2.1
 CC binding motif. These peptides bind HLA-A2.1 and have a binding
 CC affinity of at least 1% as compared to a reference peptide (AAR71293).
 CC AAR61496 has an IC50 of 3.5 and the sequence occurs at position 1807 in
 CC the HCV LORP protein. The peptides of the invention can induce
 CC cytotoxic T lymphocytes which can react with target cells. They can
 CC be used for the treatment or prophylaxis of cancer, eg. prostate
 CC cancer or lymphoma, etc.
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 55; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLFNILGGWV 10
 Db | | | | | | | |
 1 llfnilggwv 10
 RESULT 2
 AAR73104
 ID AAR73104 standard; peptide; 10 AA.
 XX
 AC AAR73104;
 XX
 DT 16-JUN-1995 (first entry)
 XX
 DE Antigen fragment 1 from HCV has binding affinity for HLA-2.1.
 XX
 KW antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
 KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
 KW human immunodeficiency virus; human papilloma virus; p53; c-ERB2;
 KW MAGE-1; melanoma antigen-1; core antigen; surface antigen;
 KW pharmaceutical composition; in vivo; ex vivo; therapeutic;
 KW diagnostic; MHC class I molecule; major histocompatibility complex;
 KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;
 KW algorithm prediction; MBP; CMV; cytomegalovirus; HSV;
 KW herpes simplex virus; influenza A; M1.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9420127-A.
 XX
 PD 15-SEP-1994.
 XX
 PF 04-MAR-1994; 94WO-US02353.
 XX
 PR 05-MAR-1993; 93US-0027146.
 PR 04-JUN-1993; 93US-0073205.
 PR 29-NOV-1993; 93US-0159184.
 XX
 PA (CYTE-) CYTEL CORP.
 XX
 PI Grey HM, Kast WM, Sette A, Sidney J;
 XX
 WPI; 1994-302678/37.
 XX
 DR Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
 PT for treatment or prophylaxis of cancer, virus infection or
 PT autoimmune diseases.
 XX
 PS Disclosure; Page 90; 138pp; English.
 XX

CC AAR73058-121 are potential peptide binders of HLA-A2.1 motif. Using
 CC motifs disclosed in the invention, these peptides were screened for
 CC further motifs. Only peptides with binding affinity of at least 1%
 CC (binding affinity is expressed as an IC50 value) as compared to the
 CC standard peptide (AAR71293) in assays. This peptide from hepatitis
 CC C virus has an binding value of 0.4400. The peptides of the invention
 CC can induce cytotoxic T lymphocytes which can react with target cells.
 CC They can be used for the treatment or prophylaxis of cancer, eg.
 CC prostate cancer or lymphoma, etc.
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 55; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLFNILGGWV 10
 Db | | | | | | | |
 1 llfnilggwv 10
 RESULT 3
 AAR78941
 ID AAR78941 standard; peptide; 10 AA.
 XX
 AC AAR78941;
 XX
 DT 01-APR-1996 (first entry)
 XX
 DE HCV NS4 1807-1817 cytotoxic T lymphocyte epitope.
 XX
 KW HCV NS4 1807-1817; cytotoxic T; CTL; epitope; helper T; HTL; cell;
 KW lymphocyte; antigens; treatment; disease prevention; hepatitis C;
 KW non-A; non-B.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9522317-A1.
 XX
 PD 24-AUG-1995.
 XX
 PF 16-FEB-1995; 95WO-US02121.
 PR 16-FEB-1994; 94US-0197484.
 XX
 PA (CYTE-) CYTEL CORP.
 XX
 PI Ceut RW, Grey H, Sette AD, Vitiello MA;
 XX
 WPI; 1995-302545/39.
 XX
 DR Compsn. inducing cytotoxic T lymphocyte response to pref. viral,
 PT bacterial, parasitic or tumour antigens - useful in the treatment
 PT and prevention of diseases associated with the antigen e.g.
 PT hepatitis B
 XX
 PS Example 12; Page 70; 109pp; English.
 XX
 CC A compsn. which induces a cytotoxic T lymphocyte (CTL) response to
 CC a hepatitis C virus (HCV) antigen (Ag) in a mammal comprises, a
 CC HCV CTL Ag response inducing peptide (i.e. AAR78941-R78955) and a
 CC lipid conjugated helper T cell inducing peptide. The compsn. is
 CC useful in the treatment and prevention of hepatitis C.
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 55; DB 16; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLFNILGGWV 10


```

Db      1 lllfnilggwv 10
      |||||
RESULT  4
AAR84576
ID      AAR84576 standard; peptide; 10 AA.
XX      AC
XX      AAR84576;
XX      DT
XX      25-APR-1996 (first entry)
XX      DE
XX      Cytotoxic T-cell epitope, aa 1807-1816 of HCV-1 NS3 region.
XX      KW
XX      Hepatitis C virus; HCV; epitope; vaccine; immunogen.
XX      OS
XX      Hepatitis C virus.
XX      PN
XX      W09525122-A1.
XX      PD
XX      21-SEP-1995.
XX      PF
XX      16-MAR-1995; 95WO-US03224.
XX      PR
XX      17-MAR-1994; 94US-0214650.
XX      PA
XX      (SCRI ) SCRIPPS RES INST.
XX      PI
XX      Cerny A, Chisari FV;
XX      DR
XX      WPI; 1995-336941/43.
XX      PT
XX      Novel molecule comprising a cytotoxic T cell epitope - used to
XX      vaccinate against hepatitis C viral infection
XX      PS
XX      Claim 1; Page 69; 85pp; English.
XX      CC
XX      AAR84570-616, AAR84885-90 and AAR91054 are all HCV-1 derived peptides
XX      from the core, E1, E2/NS1, NS2, NS3, NS4 or NS5 regions. The peptides
XX      were tested for peptide specific cytotoxic T-cell activity. The
XX      CC
XX      peptides AAR84570-77 were found to have substantial homology with a T-
XX      cell epitope and are useful in vaccines against HCV infection.
XX      SQ
XX      Sequence 10 AA;

Query Match 100.0%; Score 55; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. NO. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LLFNIIIGGWV 10
Db      1 lllfnilggwv 10
      |||||
RESULT  5
AAY10230
ID      AAY10230 standard; Peptide; 10 AA.
XX      AC
XX      AAY10230;
XX      DT
XX      12-MAY-1999 (first entry)
XX      DE
XX      T cell epitope/MHC ligand SEQ ID NO:160.
XX      KW
XX      Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
XX      immunisation; tumour; infectious disease; immunotherapy; cancer;
XX      KW
XX      malignant melanoma; viral disease; hepatitis; AIDS.
XX      OS
XX      Synthetic.
XX      OS
XX      Hepatitis C virus.
XX      PN
XX      W09902183-A2.

Query Match 100.0%; Score 55; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. NO. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LLFNIIIGGWV 10
Db      1 lllfnilggwv 10
      |||||
RESULT  6
AAY10501
ID      AAY10501 standard; Peptide; 10 AA.
XX      AC
XX      AAY10501;
XX      DT
XX      12-MAY-1999 (first entry)
XX      DE
XX      HLA Class I motif peptide SEQ ID NO:431.
XX      KW
XX      Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
XX      immunisation; tumour; infectious disease; immunotherapy; cancer;
XX      KW
XX      malignant melanoma; viral disease; hepatitis; AIDS.
XX      OS
XX      Synthetic.
XX      OS
XX      Hepatitis C virus.
XX      PN
XX      W09902183-A2.
XX      PD
XX      21-JAN-1999.
XX      PF
XX      10-JUL-1998; 98WO-US14289.
XX      PR
XX      10-DEC-1997; 97US-0988320.

```

21-JAN-1999.
 10-JUL-1998; 98WO-US14289.
 10-DEC-1997; 97US-0988320.
 10-JUL-1997; 97CA-2209815.
 (CTL-) CTL IMMUNOTHERAPIES CORP.
 Kuendig TM, Simard JJJ;
 WPI; 1999-120514/10.
 Inducing a cytotoxic T lymphocyte response - by maintaining a level
 of antigen in the lymphatic system of a mammal so as to provide a
 sustained CTL response, used to treat, e.g. AIDS
 Disclosure; Page 30; 199pp; English.
 The present invention describes a method of inducing and/or sustaining
 an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
 method comprises: (a) delivering an antigen to the mammal at a level to
 induce an immunological CTL response in the mammal; and (b) maintaining
 the level of the antigen in the mammal's lymphatic system to maintain
 the immunologic CTL response. The method can be used for the delivery of
 e.g. a differentiation antigen, a tumour-specific multilineage antigen,
 an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
 gene antigen, or a viral antigen. They can be used for the treatment of
 disease such as cancer, e.g. malignant melanoma or infectious disease,
 e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
 to the lymphatic system provides for potent CTL stimulation that takes
 place in the milieu of the lymphoid organ, and it sustains stimulation
 that is necessary to keep CTL active, cytotoxic and recirculating
 through the body. AAY10071 to AAY10639 represent examples of peptide
 antigens given in the present invention.
 Sequence 10 AA;

Query Match 100.0%; Score 55; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. NO. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLFNIIIGGWV 10
 Db 1 lllfnilggwv 10
 |||||
 RESULT 6
 AAY10501
 ID AAY10501 standard; Peptide; 10 AA.
 XX AC AAY10501;
 XX DT
 XX 12-MAY-1999 (first entry)
 XX DE
 XX HLA Class I motif peptide SEQ ID NO:431.
 XX KW
 XX Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
 XX immunisation; tumour; infectious disease; immunotherapy; cancer;
 XX KW
 XX malignant melanoma; viral disease; hepatitis; AIDS.
 XX OS
 XX Synthetic.
 XX OS
 XX Hepatitis C virus.
 XX PN
 XX W09902183-A2.
 XX PD
 XX 21-JAN-1999.
 XX PF
 XX 10-JUL-1998; 98WO-US14289.
 XX PR
 XX 10-DEC-1997; 97US-0988320.

```

PR 10-JUL-1997; 97CA-2209815.
XX (CTL1-) CTL IMMUNOTHERAPIES CORP.
PA Kuendig TM, Simard JUL;
XX WPI; 1999-120514/10.
XX
PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT of antigen in the lymphatic system of a mammal so as to provide a
PT sustained CTL response, used to treat, e.g. AIDS
XX
PS Disclosure; Page 43; 199pp; English.
XX
CC The present invention describes a method of inducing and/or sustaining
CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC method comprises: (a) delivering an antigen to the mammal at a level to
CC induce an immunological CTL response in the mammal; and (b) maintaining
CC the level of the antigen in the mammal's lymphatic system to maintain
CC the immunologic CTL response. The method can be used for the delivery of
CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC gene antigen, or a viral antigen. They can be used for the treatment of
CC disease such as cancer, e.g. malignant melanoma or infectious disease,
CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC to the lymphatic system provides for potent CTL stimulation that takes
CC place in the milieu of the lymphoid organ, and it sustains stimulation
CC that is necessary to keep CTL active, cytotoxic and recirculating
CC through the body. AAY10071 to AAY10639 represent examples of peptide
CC antigens given in the present invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
Db 1 llfnllggwv 10

RESULT 7
AAY73104
ID AAY73104 standard; Peptide; 10 AA.
XX
AC AAY73104;
XX
DT 28-FEB-2000 (first entry)
XX
DE Hepatitis C virus (HCV)-derived MHC class I (CTL) epitope, #262.
XX
KW Chimeric; pan DR epitope; expression vector;
KW promoter; major histocompatibility complex; MHC; targeting; peptide;
KW antigen; presentation; class I; cytosolic pathway;
KW endoplasmic reticulum; class II; extracellular antigen;
KW endocytic pathway; helper T lymphocyte; HTL; universal epitope;
KW cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
KW vaccine; immunity; infection; pathogen; virus; HIV; HBV; HCV;
KW hepatitis B; hepatitis C; bacterium; protozoan; tumour cell;
KW autoimmune disease; activation; antiviral; antimalarial;
KW immunoprotective.
XX
OS Synthetic.
OS Hepatitis c virus.
XX
PN WO9958658-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10646.
XX

```

```

PR 13-MAY-1998; 98US-0078904.
PR 15-MAY-1998; 98US-0085751.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes JD, Hermanson GG, Sette A, Ishioka GY, Livingston B;
PI Chesnut RW;
XX
DR WPI; 2000-039103/03.
XX
PT Expression vectors encoding major histocompatibility targeting
PT sequence, used as, e.g. tumor vaccines -
XX
PS Claim 11; Page 68; 130pp; English.
XX
CC Sequences AAY73103-Y73145 represent hepatitis C virus (HCV)-derived MHC
CC class I (CTL) epitopes which are claimed for use in the present
CC invention. The invention relates to a novel expression vector comprising
CC a promoter operably linked to a fusion gene encoding a major
CC histocompatibility complex (MHC) targeting sequence, and two or more
CC heterologous peptide epitopes. The MHC targeting sequence may be a
CC class I targeting sequence, which directs an MHC class I epitope to
CC a cytosolic pathway or to the endoplasmic reticulum, or an MHC class
CC II targeting sequence, which directs extracellular antigens to
CC enter the endocytic pathway to be processed into antigen peptides
CC for presentation on MHC class II molecules. The heterologous
CC epitopes may comprise either helper T lymphocyte (HTL) epitopes,
CC or a cytotoxic T lymphocyte (CTL) epitope and a universal HTL
CC epitope such as a pan DR epitope (PADRE). The vectors are useful
CC for stimulating an immune response in vivo, as well as for use in
CC assaying the human immunogenicity of a human T cell peptide epitope in
CC vivo in a non-human mammal. They provide a nucleic acid vaccine for
CC enhancing immunity against infectious pathogens, such as viruses (e.g.,
CC HIV, hepatitis B (HBV) and hepatitis C (HCV)) bacteria, protozoa (e.g.,
CC Plasmodium falciparum, the cause of malaria) and also tumour cells and
CC autoimmune diseases. Universal MHC class II epitopes are advantageously
CC combined with other MHC class I and class II epitopes to increase the
CC number of cells that are activated in response to a given antigen and
CC provide a broader population coverage of MHC-reactive alleles.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
Db 1 llfnllggwv 10

RESULT 8
AAJ00048
ID AAJ00048 standard; Peptide; 10 AA.
XX
AC AAJ00048;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #39.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX

```

```
PR 19-JUL-1999; 99US-0357737.
XX (EPIM-) EPIMMUNE INC.
XX
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
XX Baker DM, Cellis E, Kubo RT, Grey HM;
XX WPI; 2001-308046/32.
XX
XX A new composition useful as a vaccines against hepatitis C virus -
XX
XX Claim 4; Page 207; 214pp; English.
XX
XX The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 10 AA;
SQ

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnilggwv 10

RESULT 9
AAJ00410
ID AAJ00410 standard; Peptide; 10 AA.
XX
XX AAJ00410;
XX
XX 02-JUL-2001 (first entry)
XX Hepatitis C virus epitope #401.
XX
XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
XX antiviral.
XX
XX Hepatitis C virus.
XX WO200121189-A1.
XX
XX 29-MAR-2001.
XX
XX 19-JUL-2000; 2000WO-US19774.
XX
XX 19-JUL-1999; 99US-0357737.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
XX Baker DM, Cellis E, Kubo RT, Grey HM;
XX WPI; 2001-308046/32.
XX
XX A new composition useful as a vaccines against hepatitis C virus -
XX
XX Disclosure; Page 110; 214pp; English.
XX
XX The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 10 AA;
SQ
```

```
Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnilggwv 10

RESULT 10
AAJ01882
ID AAJ01882 standard; Peptide; 10 AA.
XX
XX AAJ01882;
XX
XX 02-JUL-2001 (first entry)
XX Hepatitis C virus epitope #1873.
XX
XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
XX antiviral.
XX
XX Hepatitis C virus.
XX WO200121189-A1.
XX
XX 29-MAR-2001.
XX
XX 19-JUL-2000; 2000WO-US19774.
XX
XX 19-JUL-1999; 99US-0357737.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
XX Baker DM, Cellis E, Kubo RT, Grey HM;
XX WPI; 2001-308046/32.
XX
XX A new composition useful as a vaccines against hepatitis C virus -
XX
XX Disclosure; Page 147; 214pp; English.
XX
XX The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
XX Sequence 10 AA;
SQ

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnilggwv 10

RESULT 11
AAJ03783
ID AAJ03783 standard; Peptide; 10 AA.
XX
XX AAJ03783;
XX
XX 02-JUL-2001 (first entry)
XX Hepatitis C virus epitope #3774.
XX
XX Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
```

```

KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Disclosure; Page 187; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnllggwv 10

RESULT 12
AAJ03862
ID AAJ03862 standard; Peptide; 10 AA.
XX
AC AAJ03862;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3853.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Disclosure; Page 187; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnllggwv 10

RESULT 13
AAJ03960
ID AAJ03960 standard; Peptide; 10 AA.
XX
AC AAJ03960;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3951.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Example 3; Page 196; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnllggwv 10

RESULT 14
```

```

PS Example 2; Page 191; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnllggwv 10

RESULT 13
AAJ03960
ID AAJ03960 standard; Peptide; 10 AA.
XX
AC AAJ03960;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #3951.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Example 3; Page 196; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 1 llfnllggwv 10

RESULT 14
```

AAJ04082
ID AAJ04082 standard; Peptide; 10 AA.
XX
AC AAJ04082;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #4073.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Example 7; Page 203; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 55; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
Db 1 lllfnllggwv 10

RESULT 15
AAJ00411
ID AAJ00411 standard; Peptide; 11 AA.
XX
AC AAJ00411;
XX
DT 02-JUL-2001 (first entry)
XX
DE Hepatitis C virus epitope #402.
XX
KW Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
KW antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200121189-A1.
XX
PD 29-MAR-2001.
XX
PF 19-JUL-2000; 2000WO-US19774.
XX
PR 19-JUL-1999; 99US-0357737.

XX (EPIM-) EPIMMUNE INC.
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-308046/32.
XX
PT A new composition useful as a vaccines against hepatitis C virus -
XX
PS Disclosure; Page 110; 214pp; English.
XX
CC The present invention describes a composition comprising a prepared
CC hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
CC These are derived from HCV HLA-binding motifs. They are useful in
CC vaccines for the prevention and treatment of HCV infection in humans. The
CC present sequence is an epitope used in the disclosure of the invention.
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 55; DB 22; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
Db 1 lllfnllggwv 10

Search completed: August 23, 2002, 10:01:01
Job time: 391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:57:53 ; Search time 61.77 Seconds
(without alignments)
3.954 Million cell updates/sec

Title: US-08-854-825-35
Perfect score: 55
Sequence: 1 LLFNILGWV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	10	1 US-08-214-650-35	Sequence 35, Appl
2	55	100.0	10	5 PCT-US95-02121-46	Sequence 46, Appl
3	55	100.0	10	5 PCT-US95-02121-125	Sequence 125, App
4	55	100.0	138	4 US-08-444-818-16	Sequence 16, Appl
5	55	100.0	177	4 US-09-263-933-25	Sequence 25, Appl
6	55	100.0	247	1 US-08-324-977-44	Sequence 44, Appl
7	55	100.0	247	2 US-08-384-616-44	Sequence 44, Appl
8	55	100.0	247	2 US-08-904-686A-44	Sequence 44, Appl
9	55	100.0	247	4 US-09-315-850-44	Sequence 44, Appl
10	55	100.0	269	3 US-09-100-557-1	Sequence 1, Appl
11	55	100.0	382	4 US-08-444-818-20	Sequence 68, Appl
12	55	100.0	460	4 US-08-444-818-68	Sequence 20, Appl
13	55	100.0	512	4 US-08-867-611-58	Sequence 58, Appl
14	55	100.0	532	4 US-08-867-611-47	Sequence 47, Appl
15	55	100.0	534	4 US-08-867-611-48	Sequence 48, Appl
16	55	100.0	597	4 US-08-867-611-16	Sequence 16, Appl
17	55	100.0	597	5 PCT-US92-06965A-21	Sequence 21, Appl
18	55	100.0	599	4 US-08-867-611-18	Sequence 18, Appl
19	55	100.0	599	5 PCT-US92-06965A-23	Sequence 23, Appl
20	55	100.0	613	4 US-08-867-611-49	Sequence 49, Appl
21	55	100.0	739	4 US-08-444-818-148	Sequence 148, App
22	55	100.0	781	4 US-08-867-611-4	Sequence 4, Appl
23	55	100.0	781	5 PCT-US92-06965A-9	Sequence 9, Appl
24	55	100.0	859	4 US-08-444-818-30	Sequence 30, Appl
25	55	100.0	971	4 US-08-867-611-52	Sequence 52, Appl
26	55	100.0	973	4 US-08-867-611-53	Sequence 53, Appl
27	55	100.0	992	4 US-08-867-611-54	Sequence 54, Appl

28	55	100.0	1021	1 US-07-910-760-12	Sequence 12, Appl
29	55	100.0	1021	1 US-08-440-519-12	Sequence 12, Appl
30	55	100.0	1021	4 US-08-440-549-12	Sequence 12, Appl
31	55	100.0	1692	4 US-09-263-933-4	Sequence 4, Appl
32	55	100.0	1692	4 US-09-263-933-11	Sequence 11, Appl
33	55	100.0	1692	4 US-09-263-933-18	Sequence 18, Appl
34	55	100.0	1786	4 US-08-444-818-54	Sequence 54, Appl
35	55	100.0	2013	1 US-08-324-977-12	Sequence 12, Appl
36	55	100.0	2013	2 US-08-384-616-12	Sequence 12, Appl
37	55	100.0	2013	2 US-08-904-686A-12	Sequence 12, Appl
38	55	100.0	2013	4 US-09-315-850-12	Sequence 12, Appl
39	55	100.0	2261	4 US-08-444-818-66	Sequence 66, Appl
40	55	100.0	2307	4 US-09-263-933-2	Sequence 2, Appl
41	55	100.0	2307	4 US-09-263-933-9	Sequence 9, Appl
42	55	100.0	2307	4 US-09-263-933-16	Sequence 16, Appl
43	55	100.0	2436	4 US-08-444-818-75	Sequence 75, Appl
44	55	100.0	2620	1 US-08-324-977-32	Sequence 32, Appl
45	55	100.0	2620	2 US-08-384-616-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-214-650-35
; Sequence 35, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silvert, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-214-650-35

Query Match 100.0%; Score 55; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLFNILGWV 10
|||||

```

Db      1 LLFNILGGWV 10
RESULT  2
PCT-US95-02121-46
; Sequence 46, Application PC/TUS9502121
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING
; NUMBER OF SEQUENCES: 153
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02121
; FILING DATE: 16-FEB-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,484
; FILING DATE: 16-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/935,811
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,491
; FILING DATE: 27-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,568
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14137-26-4PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
PCT-US95-02121-46

Query Match      100.0%; Score 55; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LLFNILGGWV 10
Db      1 LLFNILGGWV 10
RESULT  3
PCT-US95-02121-125
; Sequence 125, Application PC/TUS9502121
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING
; NUMBER OF SEQUENCES: 153
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US/08/403,590

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02121
; FILING DATE: 16-FEB-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,484
; FILING DATE: 16-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/935,811
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,491
; FILING DATE: 27-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,682
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,568
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14137-26-4PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
PCT-US95-02121-125

Query Match      100.0%; Score 55; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LLFNILGGWV 10
Db      1 LLFNILGGWV 10
RESULT  4
US-08-444-818-16
; Sequence 16, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US/08/403,590

```


; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Allisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-16

Query Match 100.0%; Score 55; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFNILGGWV 10
| | | | | | | | | |
Db 15 LFNILGGWV 24

RESULT 5
US-09-263-933-25
; Sequence 25, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0003A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 25
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: fragment of the HCV polyprotein
US-09-263-933-25

Query Match 100.0%; Score 55; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFNILGGWV 10
| | | | | | | | | |
Db 97 LFNILGGWV 106

RESULT 6
US-08-324-977-44
; Sequence 44, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC

; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeiland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-324-977-44

Query Match 100.0%; Score 55; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFNILGGWV 10
| | | | | | | | | |
Db 192 LFNILGGWV 201

RESULT 7
US-08-384-616-44
; Sequence 44, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC

```

; TITLE OF INVENTION:  CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES:  50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE:  Naughton
; STREET:  1725 K St. N.W. Suite 1000
; CITY:  Washington
; STATE:  D.C.
; COUNTRY:  U.S.A.
; ZIP:  20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
; SOFTWARE:  ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/384,616
; FILING DATE:
; CLASSIFICATION:  424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 07/769,996
; FILING DATE:  02-OCT-1991
; APPLICATION NUMBER:  JP 2-167466
; FILING DATE:  25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  JP 2-230921
; FILING DATE:  31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  JP 2-305605
; FILING DATE:  09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 07/635,451
; FILING DATE:  28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME:  Stevens-Smith, Theresa M.
; REGISTRATION NUMBER:  36,281
; REFERENCE/DOCKET NUMBER:  900703B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (202) 659-2930
; TELEFAX:  (202) 887-0357
; TELEX:  440142
; INFORMATION FOR SEQ ID NO:  44:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  247 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-08-384-616-44

```

```

Query Match      100.0%; Score 55; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  LLFNILGGWV 10
        |||||
Db      192  LLFNILGGWV 201

```

```

RESULT      8
US-08-904-686A-44
; Sequence 44, Application US/08904686A
; Patent No. 5998130
; GENERAL INFORMATION:
; APPLICANT:  OKAYAMA, Hiroto
; APPLICANT:  FUKU, Isao
; APPLICANT:  MORI, Chisato
; APPLICANT:  TAKAMIZAWA, Akahisa
; APPLICANT:  YOSHIDA, Iwao
; TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
; NUMBER OF SEQUENCES:  50
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE:  Naughton
; STREET:  1725 K St. N.W. Suite 1000
; CITY:  Washington
; STATE:  D.C.
; COUNTRY:  U.S.A.
; ZIP:  20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
; SOFTWARE:  ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/904,686A
; FILING DATE:  01-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/324,977
; FILING DATE:  18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  JP 2-167466
; FILING DATE:  25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  JP 2-230921
; FILING DATE:  31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  JP 2-305605
; FILING DATE:  09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/099,706
; FILING DATE:  30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 07/769,996
; FILING DATE:  02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 07/635,451
; FILING DATE:  28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME:  McLeland, Le-Nhung
; REGISTRATION NUMBER:  31,541
; REFERENCE/DOCKET NUMBER:  900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (202) 659-2930
; TELEFAX:  (202) 887-0357
; INFORMATION FOR SEQ ID NO:  44:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  247 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-08-904-686A-44

```

```

Query Match      100.0%; Score 55; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  LLFNILGGWV 10
        |||||
Db      192  LLFNILGGWV 201

```

```

RESULT      9
US-09-315-850-44
; Sequence 44, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT:  OKAYAMA, Hiroto
; APPLICANT:  FUKU, Isao
; APPLICANT:  MORI, Chisato
; APPLICANT:  TAKAMIZAWA, Akahisa
; APPLICANT:  YOSHIDA, Iwao
; TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
; NUMBER OF SEQUENCES:  50
; CORRESPONDENCE ADDRESS:

```

; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westernman, Hattori, McLeLand &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-330921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeLand, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-315-850-44

Query Match 100.0%; Score 55; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LFNILGGWV 10
|
Db 192 LFNILGGWV 201

RESULT 10
US-09-100-557-1
; Sequence 1, Application US/09100557
; Patent No. 6010848
; GENERAL INFORMATION:
; APPLICANT: DelVecchio, Alfred
; APPLICANT: Zhong, Weidong
; TITLE OF INVENTION: SCREENING METHODS USING AN

; TITLE OF INVENTION: ATPASE PROTEIN FROM A VIRUS OF THE FLAVIVIRIDAE FAMILY
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,557
; FILING DATE: 19-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051,582
; FILING DATE: 02-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dinner, Dara L.
; REGISTRATION NUMBER: 33,680
; REFERENCE/DOCKET NUMBER: P50675
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5017
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-100-557-1

Query Match 100.0%; Score 55; DB 3; Length 269;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LFNILGGWV 10
|
Db 96 LFNILGGWV 105

RESULT 11
US-08-444-818-68
; Sequence 68, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,590
;; FILING DATE: 14-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Harbin, Alisa A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0110.002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (508)359-3876
;; TELEFAX: (508)359-3885
;; INFORMATION FOR SEQ ID NO: 68:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 382 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-444-818-68

Query Match 100.0%; Score 55; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
Db 253 LLFNILGGWV 262

RESULT 12
US-08-444-818-20
; Sequence 20, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANV Diagnostics and vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-20

Query Match 100.0%; Score 55; DB 4; Length 460;

Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
Db 337 LLFNILGGWV 346

RESULT 13
US-08-867-611-58
; Sequence 58, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-611-58

Query Match 100.0%; Score 55; DB 4; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFNILGGWV 10

|||||

Db 378 LFNILGGWV 387

RESULT 14

US-08-867-611-47

; Sequence 47, Application US/08867611

; Patent No. 6172189

; GENERAL INFORMATION:

; APPLICANT: DEVARE, SUSHIL G

; APPLICANT: DESAI, SURESH M

; APPLICANT: CASEY, JAMES M

; APPLICANT: DAILEY, STEPHEN H

; APPLICANT: DAWSON, GEORGE J

; APPLICANT: GUTIERREZ, ROBIN A

; APPLICANT: LESNIEWSKI, RICHARD R

; APPLICANT: STEWART, JAMES L

; APPLICANT: RUPPRECHT, KEVIN R

; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT

; TITLE OF INVENTION: ANTIGENS

; NUMBER OF SEQUENCES: 59

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES

; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/867,611

; FILING DATE: 02-JUN-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/646,757

; FILING DATE:

; APPLICATION NUMBER: US/08/179,896

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/572,822

; FILING DATE: 24-AUG-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/614,069

; FILING DATE: 07-NOV-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,561

; FILING DATE: 21-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,565

; FILING DATE: 21-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,566

; FILING DATE: 21-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: POREMSKI, PRISCILLA E

; REGISTRATION NUMBER: 33,207

; REFERENCE/DOCKET NUMBER: 4834.US.P6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-6365

TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 592 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-867-611-47

Query Match 100.0%; Score 55; DB 4; Length 592;

Best Local Similarity 100.0%; Pred. No. 0.052;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFNILGGWV 10

|||||

Db 463 LFNILGGWV 472

RESULT 15

US-08-867-611-48

; Sequence 48, Application US/08867611

; Patent No. 6172189

; GENERAL INFORMATION:

; APPLICANT: DEVARE, SUSHIL G

; APPLICANT: DESAI, SURESH M

; APPLICANT: CASEY, JAMES M

; APPLICANT: DAILEY, STEPHEN H

; APPLICANT: DAWSON, GEORGE J

; APPLICANT: GUTIERREZ, ROBIN A

; APPLICANT: LESNIEWSKI, RICHARD R

; APPLICANT: STEWART, JAMES L

; APPLICANT: RUPPRECHT, KEVIN R

; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT

; TITLE OF INVENTION: ANTIGENS

; NUMBER OF SEQUENCES: 59

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES

; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/867,611

; FILING DATE: 02-JUN-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/646,757

; FILING DATE:

; APPLICATION NUMBER: US/08/179,896

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/572,822

; FILING DATE: 24-AUG-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/614,069

; FILING DATE: 07-NOV-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,561

; FILING DATE: 21-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,565

; FILING DATE: 21-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,566

; FILING DATE: 21-AUG-1991

```

; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 594 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-867-611-48

```

```

Query Match      100.0%; Score 55; DB 4; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 LFNILGGWV 10
   | | | | | | | |
Db 465 LFNILGGWV 474

```

```

Search completed: August 23, 2002, 09:57:53
Job time: 203 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:56:51 ; Search time 78.82 Seconds
(without alignments)
12.191 Million cell updates/sec

Title: US-08-854-825-35

Perfect score: 55

Sequence: 1 LLFNILGGWV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	100.0	492	2 PS0326	polyprotein - hepa
2	55	100.0	3010	1 GNVVTC	genome polyprotein
3	55	100.0	3010	1 GNVVTC	genome polyprotein
4	55	100.0	3010	1 A45573	genome polyprotein
5	55	100.0	3010	1 S18030	genome polyprotein
6	55	100.0	3010	1 GNVVTC	genome polyprotein
7	55	100.0	3011	1 GNVVTC	genome polyprotein
8	55	100.0	3011	1 GNVVTC	genome polyprotein
9	55	100.0	3011	1 S40770	genome polyprotein
10	55	100.0	3014	1 JC5620	genome polyprotein
11	44	80.0	383	2 C95186	hypothetical prote
12	44	80.0	383	2 D98053	hypothetical prote
13	44	80.0	3033	1 JQ1303	genome polyprotein
14	43	78.2	290	1 S70875	type IV prepilin p
15	43	78.2	290	2 T30215	monosaccharide tra
16	42	76.4	310	1 B65083	secretion protein
17	42	76.4	3033	1 GNVVTC	genome polyprotein
18	41	74.5	283	1 C47755	pectic enzyme secr
19	41	74.5	611	2 G83177	probable sodium/hy
20	40	72.7	131	2 H75320	hypothetical prote
21	40	72.7	359	2 AE0941	probable flmbrail
22	40	72.7	660	2 S73673	hypothetical prote
23	39	70.9	240	2 G83401	hypothetical prote
24	39	70.9	368	2 A96979	probable spore ger
25	39	70.9	409	2 T47026	hypothetical prote
26	39	70.9	409	2 AG0235	probable sugar tra
27	38	69.1	114	2 AG0666	probable membrane
28	38	69.1	181	2 E69900	hypothetical prote
29	38	69.1	294	2 D82657	glycosyl transfera

ALIGNMENTS

RESULT 1

PS0326

polyprotein - hepatitis C virus (isolate Fla) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: PS0326

R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.

Gene 105, 167-172, 1991

A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genot

A:Reference number: PS0326; MUID:92039028

A:Accession: PS0326

A:Molecule type: genomic RNA

A:Residues: 1-492 <LLJ>

A:Cross-references: GB:M60220

A:Note: This sequence corresponds to nonstructural protein NS3 region

A:Note: translation of the nucleotide sequence is not complete

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: polyprotein

Query Match 100.0%; Score 55; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10

|||||

Db 358 LLFNILGGWV 367

RESULT 2

GNVVT

genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001

C:Accession: A38465

R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.

J. Virol. 65, 1105-1113, 1991

A:Title: Structure and organization of the hepatitis C virus genome isolated from hum

A:Reference number: A38465; MUID:91140698

A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <YAK>

A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

[illegible]

F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 100.0%; Score 55; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
|||||
Db 1807 LLFNILGGWV 1816

RESULT 6
GNMWTV
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus NS3 (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A;Reference number: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein C; envelope protein M; hepatitis C virus NS3 (nonstructural protein NS4a); capsid protein C #status predicted <CPC>
F;116-191/Product: capsid protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 100.0%; Score 55; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
|||||
Db 1807 LLFNILGGWV 1816

RESULT 7
GNMWCV
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus NS3 (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:Q329873; PIDN:AAA45676.1; PID:Q329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t
A;Reference number: PQ0393; MUID:92268871
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F;116-191/Product: capsid protein C #status predicted <CPC>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 100.0%; Score 55; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
|||||
Db 1807 LLFNILGGWV 1816

RESULT 8
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus NS3 (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M. submitted to Genbank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; PIDN:AAA45534.1; PID:G329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M. Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A;Reference number: A41546; MUID:92052256
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: AP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4B #status predicted <NS4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:136, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 2681, 2682, 2683, 2684, 2685, 2686, 2687, 2688, 2689, 2690, 2691, 2692, 2693, 2694, 2695, 2696, 2697, 2698, 2699, 2700, 2701, 2702, 2703, 2704, 2705, 2706, 2707, 2708, 2709, 2710, 2711, 2712, 2713, 2714, 2715, 2716, 2717, 2718, 2719, 2720, 2721, 2722, 2723, 2724, 2725, 2726, 2727, 2728, 2729, 2730, 2731, 2732, 2733, 2734, 2735, 2736, 2737, 2738, 2739, 2740, 2741, 2742, 2743, 2744, 2745, 2746, 2747, 2748, 2749, 2750, 2751, 2752, 2753, 2754, 2755, 2756, 2757, 2758, 2759, 2760, 2761, 2762, 2763, 2764, 2765, 2766, 2767, 2768, 2769, 2770, 2771, 2772, 2773, 2774, 2775, 2776, 2777, 2778, 2779, 2780, 2781, 2782, 2783, 2784, 2785, 2786, 2787, 2788, 2789, 2790, 2791, 2792, 2793, 2794, 2795, 2796, 2797, 2798, 2799, 2800, 2801, 2802, 2803, 2804, 2805, 2806, 2807, 2808, 2809, 2810, 2811, 2812, 2813, 2814, 2815, 2816, 2817, 2818, 2819, 2820, 2821, 2822, 2823, 2824, 2825, 2826, 2827, 2828, 2829, 2830, 2831, 2832, 2833, 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843, 2844, 2845, 2846, 2847, 2848, 2849, 2850, 2851, 2852, 2853, 2854, 2855, 2856, 2857, 2858, 2859, 2860, 2861, 2862, 2863, 2864, 2865, 2866, 2867, 2868, 2869, 2870, 2871, 2872, 2873, 2874, 2875, 2876, 2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885, 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895, 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905, 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914, 2915, 2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924, 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934, 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944, 294

```
Query Match      100.0%; Score 55; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels
```

QY 1 LLENILGGWV 10
| | | | | | | |
Db 1807 LLENILGGWV 1816

Db 1807 LLFNILGGWV 1816

RESULT 9

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence.revision 19-May-2000 #text_change 19-Jan-2001

K. OKAMOTO, H. submitted to the EMBL Data Library, March 1992

Submitted to the EMBL data library, March 1992

A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OK>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>

Query Match 100.0%; Score 55; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels

Qy 1 LLENILGGWV 10
 |||||
Db 1807 LLENILGGWV 1816

Db 1807 LLENILGGWV 1816

RESULT 10

A; Residues: 1-3014 (CH₂)₁₀
A; Cross-references: GB: Y
A; Experimental source: g
A; Note: the translation

C;Keywords: ATP; glycoprotein; superfamily: nephrin-like; F;2-115/Product: capsid; F:116-191/Product: envelope

F:2-115/Product: capsid
F:116-191/Product: envelope

F; 384-408/H
F; 390-730/H

Query Match 80.0%; Score 44; DB 2; Length 383;
Best Local Similarity 70.0%; Pred. No. 3.3;
Matches 7: Conservative 3; Mismatches 0: Indels

Db 1807 LLENILGGWV 1816

QY 1 LLFNILGGWV 10
|:|:|||||:
Db 359 LVFSILGGWL 368

RESULT 12

D98053
hypothetical protein MFS transporter [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D98053
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; Eber, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98053
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-383 <R>
A:Cross-references: GB:AE007317; PIDN:AAL00257.1; PID:g15459110; GSPDB:GN00174
C:Genetics:
A:Gene: MFS transporter

Query Match 80.0%; Score 44; DB 2; Length 383;
Best Local Similarity 70.0%; Pred. No. 3.3;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
|:|:|||||:
Db 359 LVFSILGGWL 368

RESULT 13

JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: JQ1303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.; J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A:Reference number: JQ1303; MUID:92044440
A:Accession: JQ1303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 80.0%; Score 44; DB 1; Length 3033;
Best Local Similarity 70.0%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
|:|:|||||:
Db 1811 ILLNIGGWL 1820

RESULT 14

S70875
type IV prepilin peptidase (EC 3.4.99.-) tapD - Aeromonas hydrophila
N:Alternate names: prepilin type IV peptidase
C:Species: Streptococcus pneumoniae
C:Accession: D98053
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; Eber, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98053
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <PEP>
A:Cross-references: EMBL:U20255; NID:g695159; PIDN:AAC43998.1; PID:g663118
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
C:Genetics:
A:Gene: tapD

C:Superfamily: type IV prepilin peptidase
C:Keywords: hydrolase; methyltransferase; S-adenosylmethionine
Query Match 78.2%; Score 43; DB 1; Length 290;
Best Local Similarity 80.0%; Pred. No. 3.7;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
|:|:|||||:
Db 170 LLFNILGGFV 179

RESULT 15

T30215
monosaccharide transport protein - Streptomyces hygroscopicus (fragment)
C:Species: Streptomyces hygroscopicus
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30215
R:Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; St Gene 169, 9-16, 1996
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces
A:Reference number: Z20782; MUID:96186896
A:Accession: T30215
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-290 <APA>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987089; PIDN:CAA60449.1
C:Genetics:
A:Note: orfzz

Query Match 78.2%; Score 43; DB 2; Length 290;
Best Local Similarity 70.0%; Pred. No. 3.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
|:|:|||||:
Db 78 ILFNPLGGWL 87

Search completed: August 23, 2002, 09:56:53
Job time: 143 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 09:55:25 ; Search time 37 Seconds
(without alignments)
10.465 Million cell updates/sec

Title: US-08-854-825-35
Perfect score: 55
Sequence: 1 LLFNILGGWV 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	55	100.0	3010	1	POLG_HCVBK	P26663 h genome po
2	55	100.0	3010	1	POLG_HCVJA	P26662 h genome po
3	55	100.0	3010	1	POLG_HCVJT	Q00269 h genome po
4	55	100.0	3010	1	POLG_HCVTW	P29846 h genome po
5	55	100.0	3011	1	POLG_HCVI	P26664 h genome po
6	55	100.0	3011	1	POLG_HCVH	P27958 h genome po
7	44	80.0	291	1	LEP4_AERSA	O68964 aeromonas s
8	44	80.0	3033	1	POLG_HCVJ6	P26660 h genome po
9	43	78.2	290	1	LEP4_AERHY	P45794 aeromonas h
10	42	76.4	269	1	PPPA_ECOLI	Q46836 escherichia
11	42	76.4	3033	1	POLG_HCVJ8	P26661 h genome po
12	41	74.5	283	1	LEP4_ERWCH	P31711 erwinia chr
13	40	72.7	660	1	SGAT_MYCPN	P75291 mycoplasma
14	38	69.1	610	1	LKHA_HUMAN	P09960 homo sapien
15	37	67.3	130	1	FB12_MYCPN	P75450 mycoplasma
16	37	67.3	276	1	BCHL_CHLTE	Q9f714 chlorobium
17	37	67.3	597	1	ADAS_CAEEL	O45218 caenorhabdi
18	37	67.3	631	1	ADAS_DROME	Q9v778 drosophila
19	36	65.5	216	1	CLPP_MATZE	P26567 zea mays (m
20	36	65.5	456	1	UHPT_CHLMU	Q9pjl8 chlamydia m
21	36	65.5	3175	1	RPOA_EAV	P19811 equine arte
22	35	63.6	178	1	VNCA_RSYM	Q01209 rice stripe
23	35	63.6	178	1	VNCA_RSVT	Q00844 rice stripe
24	35	63.6	181	1	COTE_BACSU	P14016 escherichia
25	35	63.6	402	1	OPDE_PSEAE	Q01602 pseudomonas
26	35	63.6	457	1	YG90_HREIN	P45320 haemophilus
27	35	63.6	658	1	ADAS_HUMAN	O00116 homo sapien
28	35	63.6	804	1	YBPP_ECOLI	P77504 escherichia
29	35	63.6	989	1	YD30_YEAST	P32898 saccharomyc
30	35	63.6	1926	1	LPH_RABIT	P09849 oryctolagus
31	35	63.6	1927	1	LPH_HUMAN	P09848 homo sapien
32	35	63.6	1928	1	LPH_RAT	Q02401 rattus norv
33	34	61.8	173	1	CRGF_BOVIN	P23005 bos taurus

RESULT 1

ID	POLG_HCVBK	STANDARD;	PRT;	3010 AA.
AC	P26663;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate BK) (HCV).			
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus			
OC	Hepacivirus			
OX	NCBI_Taxid=11105;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91140698; PubMed=1847440;			
RA	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;			
RT	"Structure and organization of the hepatitis C virus genome isolated from human carriers.";			
RL	J. Virol. 65:1105-1113(1991).			
RN	[2]			
RP	SEQUENCE OF 1487-1500.			
RX	MEDLINE=96235224; PubMed=8647104;			
RA	Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;			
RT	"Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase.";			
RL	Eur. J. Biochem. 237:611-618(1996).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.			
RX	MEDLINE=97015088; PubMed=8861916;			
RA	Love R.A., Farge H.E., Wickersham J.A., Hostomsky Z., Habuka N., Moomaw E.W., Adachi T., Hostomska Z.;			
RT	"The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site.";			
RL	Cell 87:331-342(1996).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.			
RX	MEDLINE=98227846; PubMed=9568891;			
RA	Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;			
RT	"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form.";			
RL	Protein Sci. 7:837-847(1998).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A			

Q92z43 squalus aca
P72640 synechocyst
Q58578 methanococc
P28246 escherichia
Q9eqd2 rattus norv
Q927n9 chlamydia p
P06818 influenza a
O91744 influenza a
O91745 influenza a
P03473 influenza a
P03484 influenza a
P06820 influenza a

ALIGNMENTS

LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND MRNA.

-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M58335; AAA72945.1; -

PIR: A38465; GNMVTC.

PDB: 1AIQ; 25-MAR-98.

PDB: LJXP; 14-JAN-98.

PDB: 1NS3; 08-APR-98.

MEROPS: S29.001; -

MEROPS: U39.001; -

InterPro: IPR001410; DEAD.

InterPro: IPR002531; HCV_NSI.

InterPro: IPR002518; HCV_NS2.

InterPro: IPR004109; HCV_NS3.

InterPro: IPR000745; HCV_NS4a.

InterPro: IPR001490; HCV_NS4b.

InterPro: IPR002868; HCV_NS5a.

InterPro: IPR002166; HCV_RDRP.

InterPro: IPR002522; HCV_capsid.

InterPro: IPR002521; HCV_core.

InterPro: IPR002519; HCV_env.

InterPro: IPR001650; Helicase_C.

Pfam: PF01543; HCV_capsid; 1.

Pfam: PF01542; HCV_core; 1.

Pfam: PF01539; HCV_env; 1.

Pfam: PF01560; HCV_NSI; 1.

Pfam: PF01538; HCV_NS2; 1.

Pfam: PF02907; HCV_NS3; 1.

Pfam: PF01006; HCV_NS4a; 1.

Pfam: PF01001; HCV_NS4b; 1.

Pfam: PF01506; HCV_NS5a; 1.

Pfam: PF00998; HCV_RDRP; 1.

Pfam: PF00271; helicase_C; 1.

ProDom: PD186062; HCV_NSI; 1.

PolyProtein: Glycoprotein; Transferrase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).

CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

CHAIN 192 383 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

CHAIN 384 729 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

CHAIN 730 1006 PROTEASE/HELICASE NS3 (POTENTIAL).

CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).

CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

CHAIN 1863 2013 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

CHAIN 3010 347 CHARGE RELAY SYSTEM.

CHAIN 347 369 CHARGE RELAY SYSTEM.

ACT_SITE 1083 1107 CHARGE RELAY SYSTEM.

ACT_SITE 1107 1165 CHARGE RELAY SYSTEM.

ACT_SITE 1165 1237 ATP (POTENTIAL).

NP_BIND 1230 1237 DECH_BOX.

SITE 1316 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 136 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2798 2798 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDF9C CRC64;

Query Match 100.0%; Score 55; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10

Db 1807 LLFNILGGWV 1816

RESULT 2

POLG_HCVJA STANDARD; PRT; 3010 AA.

AC P26662;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Japanese) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI_TaxID=11116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91088550; PubMed=2175903;

RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;

RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.;"

RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).

RN [2]

RP DISCUSSION OF SEQUENCE.

RX MEDLINE=91192160; PubMed=1849488;

RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K., Ohkoshi S., Shimotohno K.;

RT "Molecular structure of the Japanese hepatitis C viral genome.;"

RL FEBS Lett. 280:325-328(1991).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

DR InterPro: IPR002519; HCV_env.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR SMART: SM00492; HELIC3; 1.
 DR ProDom: PD186062; HELIC3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 326573 MW; 94AIC77435D642BB CRC64;
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1093
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT ACT_SITE 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94AIC77435D642BB CRC64;
 Query Match 100.0%; Score 55; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LFNILGGWV 10
 |||||
 Db 1807 LFNILGGWV 1816
 RESULT 4
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein {Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)}.
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OC NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M84754; -; NOT_ANNOTATED_CDS.
 CC PIR; A40244; GNVVTV.
 CC HSSP; P26663; LJXP.
 CC MEROPS; S29.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRP; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 326573 MW; 94AIC77435D642BB CRC64;
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1093
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT ACT_SITE 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94AIC77435D642BB CRC64;


```

FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1165 1237 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 100.0%; Score 55; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLFNILGGWV 10
    |||||
Db 1807 LLFNILGGWV 1816

RESULT 5
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (BC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (BC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (BC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC. SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M62321; AAA45676.1; -.
CC PIR: A39166; GNWVC3.
CC HSPSP; P27958; IHEI.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRp.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00998; HCV_RdRp; 1.
CC Pfam; PF00271; Helicase_C; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
CC SMART; SM00492; HELIC3; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
```

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 55; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFNILGGWV 10
 |||||
 Db 1807 LFNILGGWV 1816

RESULT 6
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=158800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RA "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 [2]
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 [3]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF

NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 ACTIVATION OF NS3.
 -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M67463; AAA45534.1; -.
 DR PIR; A36814; GMYVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; HCV_RDRP; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00492; HELICC3; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 FT 3D-structure. 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 INIT_MET CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 191 CAPSID PROTEIN C.
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 747 809 PROTEIN P7.
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 FT CHAIN 347 369 POTENTIAL.
 FT TRANSMEM

FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 ATP (POTENTIAL).
 FT SITE 1316 DECH BOX.
 FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CDD94753 CRC64;

 Query Match 100.0%; Score 55; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 LFNILGGWV 10
 Db 1807 LFNILGGWV 1816

 RESULT 7
 LEP4_AERSA
 ID LEP4_AERSA STANDARD; PRT; 291 AA.
 AC O68964; O54483;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Type 4 prelin-like proteins leader peptide processing enzyme
 DE [includes: Leader peptidase (EC 3.4.99.-) (Prelin peptidase); N-
 DE methyltransferase (EC 2.1.1.-)].
 GN TAPD OR PILD.
 OS Aeromonas salmonicida.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OX NCBI_TaxID=645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A450;
 RA Pepe C.M., Strom M.S.;
 RT "Aeromonas salmonicida type IV prelin peptidase and type IV pilus
 RT assembly genes".
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A449;
 RA Lutwyche P., Perez-Casal J.F., Crump E.M., Trust T.J.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CLEAVES TYPE-4 FIBRILLAR LEADER SEQUENCE AND METHYLATES
 CC THE N-TERMINAL (GENERALLY PHE) RESIDUE. PROCESSES THE TAPA PILIN
 CC PRECURSOR DURING MEMBRANE TRANSLOCATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24; ALSO KNOWN AS TYPE
 CC III LEADER PEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC
 DR EMBL; AF059249; AAC23569.1; -
 DR EMBL; U95640; AAC04561.1; -
 DR MEROPS; A24.001; -
 DR InterPro; IPR000045; Peptidase_C20.
 DR Pfam; PF01478; Peptidase_C20; 1.
 DR PRINTS; PR00864; PREPILNPTASE.
 KW Multifunctional enzyme; Hydrolase; Protease; Transferase;
 KW Methyltransferase; Transmembrane; Inner membrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 FT CONFLICT 60 80 P -> A (IN REF. 2).
 FT CONFLICT 84 84 A -> G (IN REF. 2).
 FT CONFLICT 100 100 C -> W (IN REF. 2).
 FT CONFLICT 103 103 C -> W (IN REF. 2).
 FT CONFLICT 126 126 A -> G (IN REF. 2).
 FT CONFLICT 163 163 T -> N (IN REF. 2).
 FT CONFLICT 198 198 S -> R (IN REF. 2).
 FT CONFLICT 253 253 L -> V (IN REF. 2).
 SQ SEQUENCE 291 AA; 32500 MW; 4CFFA0AFC886E368 CRC64;

 Query Match 80.0%; Score 44; DB 1; Length 291;
 Best Local Similarity 80.0%; Pred. No. 1.4;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 1 LFNILGGWV 10
 Db 171 LFNILGGWV 180

 RESULT 8
 POLG_HCVJ6
 ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions".
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with asp or Glu in the p6
 CC position, Cys or Thr in E1 and Ser or Ala in P1.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D00944; BAA00792.1; -
 CC PIR: JQ1303; JQ1303.
 CC HSP: P27938; IHEI.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002538; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; HCV_RdRP; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC ProbDom: PD186062; HCV_NS1; 1.
 CC PolyProtein: Glycoprotein; Transferase: RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 CC INIT_MET 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 3033 347
 CC TRANSMEM 347 369
 CC ACT_SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 1323 1323
 CC CARBOHYD 136 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578

FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 80.0%; Score 44; DB 1; Length 3033;
 Best Local Similarity 70.0%; Pred. No. 12;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLFNILGGWV 10
 DB 1811 ILLNILGGWL 1820
 RESULT 9
 LEP4_AERHY
 ID LEP4_AERHY STANDARD; PRT; 290 AA.
 AC P45794;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Type 4 prepilin-like proteins leader peptide processing enzyme
 DE [Includes: Leader peptidase (EC 3.4.99.-) (Prepilin peptidase); N-
 DE methyltransferase (EC 2.1.1.-)].
 GN TAPD.
 OS Aeromonas hydrophila.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OX NCBI_TaxID=644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AH65;
 RX MEDLINE=96417863; PubMed=8820654;
 RA Pepe C.M., Eklund M.W., Strom M.S.;
 RT "Cloning of an Aeromonas hydrophila type IV pilus biogenesis gene
 RT cluster: complementation of pilus assembly functions and
 RT characterization of a type IV leader peptidase/N-methyltransferase
 RT required for extracellular protein secretion.";
 RL Mol. Microbiol. 19:857-869(1996)
 CC -!- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES
 CC THE N-TERMINAL (GENERALLY PHE) RESIDUE. PROCESSES THE TAPA PILIN
 CC PRECURSOR DURING MEMBRANE TRANSLOCATION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24; ALSO KNOWN AS TYPE
 CC III LEADER PEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U20255; AAC43998.1; -
 CC MEROPS: A24.001; -
 CC InterPro: IPR000045; Peptidase_C20.
 CC Pfam: PF01478; Peptidase_C20; 1.
 CC PRINTS: PR00864; PREPILNPASE.
 CC Multifunctional enzyme; Hydrolase; Protease; Transferase;
 KW Methyltransferase; Transmembrane; Inner membrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 130 150 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 261 281 POTENTIAL.
 SQ SEQUENCE 290 AA; 32308 MW; EA584F277041A99B CRC64;

```

Query Match          78.2%; Score 43; DB 1; Length 290;
Best Local Similarity 80.0%; Pred. No. 2.1;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
Db 170 LLENLLGGFV 179
||||:||||

RESULT 10
PPPA_ECOLI          STANDARD; PRT; 269 AA.
AC Q46836;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leader peptidase pppa.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24; ALSO KNOWN AS TYPE
CC III LEADER PEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28377; AAA69139.1; ALT_INIT.
CC EMBL; AE000379; AAC76008.1; ALT_INIT.
CC EcoGene; EG12992; pppa.
DR InterPro: IPR000045; Peptidase_C20.
DR Pfam; PF01478; Peptidase_C20; 1.
DR PRINTS; PR00864; PREPILNPTASE.
KW Complete proteome.
SQ SEQUENCE 269 AA; 29466 MW; E793FD8B4C81F6A4 CRC64;

Query Match          76.4%; Score 42; DB 1; Length 269;
Best Local Similarity 80.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
Db 210 LLFAALGGWV 219
||| |||||

RESULT 11
POLG_HCVJ8          STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -! FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -! CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -! SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -! SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10988; BAA01761.1; -
CC PIR; A40250; GNMVJ8.
DR HSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; HCV_RdRP; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.

```

```

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 76.4%; Score 42; DB 1; Length 3033;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLEFNLGGWV 10
DB 1811 ILLNMGWL 1820

RESULT 12
LEP4_ERWCH
AC P31711; STANDARD; PRT; 283 AA.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type 4 prelin-like proteins leader peptide processing enzyme (Pectic
DE enzymes secretion protein outo) [includes: leader peptidase
DE (EC 3.4.99.-) (Prelinin peptidase); N-methyltransferase (EC 2.1.1.-)].
GN OUTO.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=93054355; PubMed=1429461;
RA Lindeberg M., Collmer A.;
RT "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other Gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
CC -!- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES
CC THE N-TERMINAL (GENERALLY PHE) RESIDUE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

```

```

CC (Probable).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24; ALSO KNOWN AS TYPE
CC III LEADER PEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L02214; AAA24841.1; .
CC DR PIR; C47755; C47755.
CC DR MEROPS; A24.001; .
CC DR InterPro; IPR000045; Peptidase_C20.
CC DR Pfam; PF01478; Peptidase_C20; 1.
CC DR PRINTS; PR00864; PREPILNPTASE.
CC KW Multifunctional enzyme; Hydrolase; Protease; Transferrase;
CC Methyltransferase; Transmembrane; Inner membrane.
CC FT TRANSMEM 13 33 POTENTIAL.
CC FT TRANSMEM 106 126 POTENTIAL.
CC FT TRANSMEM 128 148 POTENTIAL.
CC FT TRANSMEM 153 173 POTENTIAL.
CC FT TRANSMEM 176 196 POTENTIAL.
CC FT TRANSMEM 216 236 POTENTIAL.
CC FT TRANSMEM 259 279 POTENTIAL.
CC SQ SEQUENCE 283 AA; 31355 MW; 90B9CE722C4AAA7E CRC64;

Query Match 74.5%; Score 41; DB 1; Length 283;
Best Local Similarity 66.7%; Pred. No. 4.5;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLEFNLGGW 9
DB 273 LVLNLGGW 281

RESULT 13
SGAT_MYCPN
ID SGAT_MYCPN STANDARD; PRT; 660 AA.
AC P75291;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative transport protein sgat homolog.
DE SGAT OR MPN496 OR MF347.
GN Mycoplasma pneumoniae.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: COULD ACT AS THE TRANSPORT PROTEIN FOR THE UNKNOWN
CC PENITOL SUBSTRATE OF THE SGA OPERON (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC EMBL; AE000033; AAB95995.1; -
DR TransPort; Transmembrane; Complete proteome.
KW TRANSSEM 14 34 POTENTIAL.
FT TRANSSEM 74 94 POTENTIAL.
FT TRANSSEM 98 118 POTENTIAL.
FT TRANSSEM 162 182 POTENTIAL.
FT TRANSSEM 194 214 POTENTIAL.
FT TRANSSEM 224 244 POTENTIAL.
FT TRANSSEM 307 327 POTENTIAL.
FT TRANSSEM 362 382 POTENTIAL.
FT TRANSSEM 401 421 POTENTIAL.
FT TRANSSEM 422 442 POTENTIAL.
FT TRANSSEM 450 470 POTENTIAL.
FT TRANSSEM 480 500 POTENTIAL.
FT TRANSSEM 592 612 POTENTIAL.
FT TRANSSEM 615 635 POTENTIAL.
SQ SEQUENCE 660 AA; 70858 MW; AA45EC672921BAA8 CRC64;

Query Match 72.7% Score 40; DB 1; Length 660;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FNIIGGW 10
Db 612 YNIGGW 619

RESULT 14
LKHA_HUMAN STANDARD; PRT; 610 AA.
AC P09960;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Leukotriene A-4 hydrolase (EC 3.3.2.6) (LTA-4 hydrolase) (Leukotriene
DE A(4) hydrolase).
GN LTA4H OR LTA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007621; PubMed=3654641;
RA Minami M., Ohno S., Kawasaki H., Raedmark O., Samuelsson B.,
RA Joernvall H., Shimizu T., Seyama Y., Suzuki K.;
RT "Molecular cloning of a cDNA coding for human leukotriene A4
RT hydrolase. Complete primary structure of an enzyme involved in
RT eicosanoid synthesis.";
RL J. Biol. Chem. 262:13873-13876(1987).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88016154; PubMed=2821541;
RA Funk C.D., Raedmark O., Fu J.Y., Matsumoto T., Joernvall H.,
RA Shimizu T., Samuelsson B.;
RT "Molecular cloning and amino acid sequence of leukotriene A4
RT hydrolase.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6677-6681(1987).
RN [3]
RP ZINC-BINDING, PEPTIDASE ACTIVITY, AND SIMILARITY TO ZINC PROTEASES.
RX MEDLINE=90365706; PubMed=1975494;
RA Toh H., Minami M., Shimizu T.;
RT "Molecular evolution and zinc ion binding motif of leukotriene A4
RT hydrolase.";
RL Biochem. Biophys. Res. Commun. 171:216-221(1990).
RN [4]
RP ZINC-BINDING, AND PEPTIDASE ACTIVITY.
RX MEDLINE=91058588; PubMed=2244921;
RA Haeggstroem J.Z., Wetterholm A., Shapiro R., Vallee B.L.,
RA Samuelsson B.;
RT "Leukotriene A4 hydrolase: a zinc metalloenzyme.";
RL Biochem. Biophys. Res. Commun. 172:965-970(1990).
RN [5]
RP MUTAGENESIS OF ZINC-LIGANDS.
RX MEDLINE=91352040; PubMed=1881903;
RA Medina J.F., Wetterholm A., Raedmark O., Shapiro R., Haeggstroem J.Z.,
RA Vallee B.L., Samuelsson B.;
RT "Leukotriene A4 hydrolase: determination of the three zinc-binding
RT ligands by site-directed mutagenesis and zinc analysis.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:7620-7624(1991).
RN [6]
RP MUTAGENESIS OF GLU-296.
RX MEDLINE=92387388; PubMed=1516710;
RA Minami M., Bito H., Ohishi N., Tsuge H., Miyano M., Mori M.,
RA Wada H., Mutoh H., Shimada S., Izumi T., Abe K., Shimizu T.;
RT "Leukotriene A4 hydrolase, a bifunctional enzyme. Distinction of
RT leukotriene A4 hydrolase and aminopeptidase activities by
RT site-directed mutagenesis at Glu-297.";
RL FEBS Lett. 309:353-357(1992).
RN [7]
RP MUTAGENESIS OF GLU-296.
RX MEDLINE=93028420; PubMed=1357660;
RA Wetterholm A., Medina J.F., Raedmark O., Shapiro R., Haeggstroem J.Z.,
RA Vallee B.L., Samuelsson B.;
RT "Leukotriene A4 hydrolase: abrogation of the peptidase activity by
RT mutation of glutamic acid-296.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9141-9145(1992).
CC -1- FUNCTION: HYDROLYSES AN EPOXIDE MOIETY OF LTA-4 TO FORM LTB-4.
CC THE ENZYME ALSO HAS SOME PEPTIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: (7E,9E,11Z,14Z)-(5S,6S)-5,6-epoxycosa-
CC 7,9,11,14-tetraenoate + H(2)O = (6Z,8E,10E,14Z)-(5S,12R)-5,12-
CC dihydroxyicosa-6,8,10,14-tetraenoate.
CC -1- COFACTOR: BINDS 1 ZINC ION.
CC -1- PATHWAY: THIRD STEP IN THE BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U27293; AAA89077.1; -
DR EMBL; U27275; AAA89077.1; JOINED.
DR EMBL; U27276; AAA89077.1; JOINED.
DR EMBL; U27277; AAA89077.1; JOINED.
DR EMBL; U27278; AAA89077.1; JOINED.
DR EMBL; U27279; AAA89077.1; JOINED.
DR EMBL; U27280; AAA89077.1; JOINED.
DR EMBL; U27281; AAA89077.1; JOINED.
DR EMBL; U27282; AAA89077.1; JOINED.
DR EMBL; U27283; AAA89077.1; JOINED.
DR EMBL; U27284; AAA89077.1; JOINED.
DR EMBL; U27285; AAA89077.1; JOINED.
DR EMBL; U27286; AAA89077.1; JOINED.
DR EMBL; U27287; AAA89077.1; JOINED.
DR EMBL; U27288; AAA89077.1; JOINED.
DR EMBL; U27289; AAA89077.1; JOINED.
DR EMBL; U27290; AAA89077.1; JOINED.
DR EMBL; U27291; AAA89077.1; JOINED.
DR EMBL; U27292; AAA89077.1; JOINED.
DR EMBL; J03459; AAA36176.1; -
DR EMBL; J02959; AAA36177.1; -
DR PIR; A27415; A27415.
DR PIR; A33886; A33886.
DR MEROPS; M01.004; -.
DR MIM; 151570; -.
DR InterPro; IPR001930; Aladiptase.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
```

DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Multifunctional enzyme; Hydrolase; Leukotriene biosynthesis; zinc;
FT INIT_MET 0 0
FT BINDING 199 199 SUBSTRATE (POTENTIAL).
FT METAL 295 295 ZINC (CATALYTIC).
FT ACT_SITE 296 296
FT METAL 299 299 ZINC (CATALYTIC).
FT METAL 318 318 ZINC (CATALYTIC).
FT ACT_SITE 383 383 PROTON DONOR (POTENTIAL).
FT MUTAGEN 295 295 H->Y: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 296 296 E->Q: LOOSE AMINOPEPTIDASE ACTIVITY, BUT
KEEPS LTA4 ACTIVITY.
FT MUTAGEN 296 296 E->A: LOOSE BOTH ACTIVITIES.
FT MUTAGEN 296 296 E->K: LOOSE BOTH ACTIVITIES.
FT MUTAGEN 299 299 H->L: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 318 318 E->A: COMPLETE LOSS OF ACTIVITY.
SQ SEQUENCE 610 AA; 69154 MW; 5A69E8505EC3382 CRC64;

Query Match 69.1%; Score 38; DB 1; Length 610;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FNILGWM 9
DB 340 FNALGGW 346
|| ||||

RESULT 15
YB12_MYCPN STANDARD; PRT; 130 AA.
AC P75450;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN112 (C09_orf130b).
GN MPN112 OR MP042.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000005; AAB95690.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 73 93
FT TRANSMEM 107 127 POTENTIAL.
SQ SEQUENCE 130 AA; 14729 MW; CCF7F807EB9968D7 CRC64;

Query Match 67.3%; Score 37; DB 1; Length 130;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 83 LLFYIPGGWL 92
||| | |||

Search completed: August 23, 2002, 09:55:27
Job time: 57 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	Match				
1	55	100.0	41	12	O11845	hepatitis c	
2	55	100.0	44	12	P87842	hepatitis c	
3	55	100.0	44	12	P87843	hepatitis c	
4	55	100.0	44	12	P87844	hepatitis c	
5	55	100.0	44	12	P87845	hepatitis c	
6	55	100.0	44	12	P87870	hepatitis c	
7	55	100.0	44	12	P87871	hepatitis c	
8	55	100.0	44	12	P87872	hepatitis c	
9	55	100.0	44	12	P90404	hepatitis c	
10	55	100.0	44	12	P87874	hepatitis c	
11	55	100.0	44	12	P90405	hepatitis c	
12	55	100.0	44	12	P90406	hepatitis c	
13	55	100.0	44	12	P90407	hepatitis c	
14	55	100.0	44	12	P87875	hepatitis c	
15	55	100.0	44	12	P90408	hepatitis c	
16	55	100.0	44	12	P90409	hepatitis c	

```
ID P87842 PRELIMINARY; PRT; 44 AA.
AC P87842;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FACTOR 8 CONCENTRATE 6681;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284290; CAB06342.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 44
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4387 MW; 79FD55078233F79E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 18 LLFNILGGWV 27

RESULT 3
P87843
ID P87843 PRELIMINARY; PRT; 44 AA.
AC P87843;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FACTOR 8 CONCENTRATE 6681;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284291; CAB06343.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 44
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4387 MW; 79FD55078233F79E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 18 LLFNILGGWV 27

RESULT 4
P87844
ID P87844 PRELIMINARY; PRT; 44 AA.
AC P87844;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FACTOR 8 CONCENTRATE 6681;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284292; CAB06344.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 44
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4357 MW; 79FD551A34E8879E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 18 LLFNILGGWV 27

RESULT 5
P87845
ID P87845 PRELIMINARY; PRT; 44 AA.
AC P87845;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FACTOR 8 CONCENTRATE 6681;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284293; CAB06345.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 44
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4387 MW; 79FD55078233F79E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLFNILGGWV 10
DB 18 LLFNILGGWV 27
```

```

|||||
18 LFNILGGWV 27

RESULT 6
P87870
ID P87870 PRELIMINARY; PRT; 44 AA.
AC P87870;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1982;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; Z84319; CAB06371.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4305 MW; 79EC1107981B550E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFNILGGWV 10
Db 18 LFNILGGWV 27

RESULT 7
P87871
ID P87871 PRELIMINARY; PRT; 44 AA.
AC P87871;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1982;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; Z84320; CAB06372.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4305 MW; 79EC1107981B550E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFNILGGWV 10
Db 18 LFNILGGWV 27

RESULT 8
P87872
ID P87872 PRELIMINARY; PRT; 44 AA.
AC P87872;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1982;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; Z84321; CAB06373.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4305 MW; 79EC1107981B550E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFNILGGWV 10
Db 18 LFNILGGWV 27

RESULT 9
P90404
ID P90404 PRELIMINARY; PRT; 44 AA.
AC P90404;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1983;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; Z84323; CAB06375.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4351 MW; 79FD55078B3A670E CRC64;
```

```
Query Match      100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
   |||||
Db 18 LLENILGGWV 27

RESULT 10
P87874
ID P87874 PRELIMINARY; PRT; 44 AA.
AC P87874;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1983;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284324; CAB06376.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 44
FT NON_TER 1
SQ SEQUENCE 44 AA; 4321 MW; 79FD491A4AFA670E CRC64;

Query Match      100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
   |||||
Db 18 LLENILGGWV 27

RESULT 11
P90405
ID P90405 PRELIMINARY; PRT; 44 AA.
AC P90405;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1983;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284325; CAB06377.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 44
FT NON_TER 1
SQ SEQUENCE 44 AA; 4321 MW; 79FD491A4AFA670E CRC64;
```

```
FT NON_TER 44
SQ SEQUENCE 44 AA; 4351 MW; 79FD55078B3A670E CRC64;

Query Match      100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
   |||||
Db 18 LLENILGGWV 27

RESULT 12
P90406
ID P90406 PRELIMINARY; PRT; 44 AA.
AC P90406;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1983;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284326; CAB06378.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 44
FT NON_TER 1
SQ SEQUENCE 44 AA; 4351 MW; 79FD55078B3A670E CRC64;

Query Match      100.0%; Score 55; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLENILGGWV 10
   |||||
Db 18 LLENILGGWV 27

RESULT 13
P90407
ID P90407 PRELIMINARY; PRT; 44 AA.
AC P90407;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NS4 PROTEIN (FRAGMENT).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 1, 1984;
RX MEDLINE=97201623; PubMed=9049409;
RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
e epitope of hepatitis C virus in two haemophilic patients.";
RL J. Gen. Virol. 78:583-590(1997).
DR EMBL; 284327; CAB06379.1; -.
```

DR InterPro; IPR000745; HCV_NS4a.
 DR Pfam; PF01006; HCV_NS4a; 1.
 FT NON_TER 1
 FT NON_TER 44 44
 SQ SEQUENCE 44 AA; 4351 MW; 79FD55078B3A670E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
 Best Local Similarity 100.0%; Pred. No. 0.0085;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
 Db 18 LLFNILGGWV 27
 |||||

RESULT 14

P87875
 ID P87875 PRELIMINARY; PRT; 44 AA.
 AC P87875;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE NS4 PROTEIN (FRAGMENT).
 GN NS4.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11103;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=PATIENT 1, 1984;
 RX MEDLINE=97201623; PubMed=9049409;
 RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
 RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL
 e epitope of hepatitis C virus in two haemophilic patients.";
 RL J. Gen. Virol. 78:583-590(1997).
 DR EMBL; Z84328; CAB06380.1;
 DR InterPro; IPR000745; HCV_NS4a.
 DR Pfam; PF01006; HCV_NS4a; 1.
 FT NON_TER 1
 FT NON_TER 44 44
 SQ SEQUENCE 44 AA; 4381 MW; 79FD55078B26170E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
 Best Local Similarity 100.0%; Pred. No. 0.0085;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
 Db 18 LLFNILGGWV 27
 |||||

RESULT 15

P90408
 ID P90408 PRELIMINARY; PRT; 44 AA.
 AC P90408;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE NS4 PROTEIN (FRAGMENT).
 GN NS4.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=PATIENT 1, 1984;
 RX MEDLINE=97201623; PubMed=9049409;
 RA Devereux H.L., Brown D., Dusheiko G.M., Emery V.C., Lee C.A.;
 RT "Long-term evolution of the 5'UTR and a region of NS4 containing a CTL

RT e epitope of hepatitis C virus in two haemophilic patients.";
 RL J. Gen. Virol. 78:583-590(1997).
 DR EMBL; Z84329; CAB06381.1;
 DR InterPro; IPR000745; HCV_NS4a.
 DR Pfam; PF01006; HCV_NS4a; 1.
 FT NON_TER 1
 FT NON_TER 44 44
 SQ SEQUENCE 44 AA; 4351 MW; 79FD55078B3A670E CRC64;

Query Match 100.0%; Score 55; DB 12; Length 44;
 Best Local Similarity 100.0%; Pred. No. 0.0085;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLFNILGGWV 10
 Db 18 LLFNILGGWV 27
 |||||

Search completed: August 23, 2002, 10:22:01
 Job time: 1501 sec

